

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:15:10 ; Search time 1741.97 Seconds
(without alignments)
10198.307 Million cell updates/sec

Title: US-10-077-130-6_COPY_19500_22500
Perfect score: 3001
Sequence: 1 cctacgagcagaactcggg.....ccacctggagcaaacgga 3001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3001	100.0	4936	AAH46904	Aah46904 cDNA enco
2	3001	100.0	8106	ABX11641	Abx11641 Human ser
3	3001	100.0	24120	ABX11642	Abx11642 Human ser
4	2999.4	99.9	5007	AAC62286	Aac62286 cDNA enco
5	2999.4	99.9	7928	AAC62287	Aac62287 cDNA enco
6	2997.8	99.9	5207	AAI43909	Aai43909 Human kin
7	2997.8	99.9	5207	AAI43908	Aai43908 Human kin
8	2996.2	99.8	5454	AAD38865	Aad38865 Human kin
9	2643.4	88.1	4175	AAC62285	Aac62285 cDNA enco
10	1706.8	56.9	3225	AAD26467	Aad26467 Human kin
11	223.2	7.4	7710	AAF44662	Aaf44662 Novel pro
12	223.2	7.4	7710	ADI29360	Adi29360 Human MAR
13	223.2	7.4	7789	AD303565	Ad303565 Human kin
14	223.2	7.4	9698	ADE47675	Ade47675 Human NOV
15	223.2	7.4	9698	ADJ78945	Adj78945 Human NOV
16	223.2	7.4	9807	ADB79958	Adb79958 Human kin
17	223.2	7.4	10662	ADC99127	Adc99127 Human KPP
18	204	6.8	9930	ADE47671	Ade47671 Human NOV
19	204	6.8	9930	ADJ78941	Adj78941 Human NOV
20	204	6.8	10122	ADE47673	Ade47673 Human NOV

21	204	6.8	10122	12	ADJ78943	Adj78943 Human NOV
22	173.6	5.8	860	10	ADE47677	Ade47677 Human NOV
23	173.6	5.8	860	12	ADJ78947	Adj78947 Human NOV
24	172	5.7	1429	2	AAX34657	Aax34657 Murine ZI
25	170.2	5.7	1365	2	ADO71726	Ado71726 Human gen
26	170.2	5.7	2055	13	ADN60271	Adn60271 Human ZIP
27	170.2	5.7	2055	13	ADR83403	Adr83403 Human ZIP
28	170.2	5.7	2079	4	AAK94258	Aak94258 Human ful
29	170.2	5.7	2079	12	ADL30841	Adl30841 Full leng
30	170.2	5.7	2079	13	ADQ68012	Adq68012 Recombina
31	170.2	5.7	2104	12	ADQ84258	Adq84258 Human tum
32	170.2	5.7	2104	13	ADQ85644	Adq85644 Human tum
33	170.2	5.7	2104	13	ADQ86740	Adq86740 Human tum
34	170.2	5.7	2226	10	ADD29770	Add29770 Human tum
35	168.6	5.6	2105	6	ABV72291	Abv72291 Nucleotid
36	168.6	5.6	2105	13	ADN60267	Adn60267 Human dea
37	168.6	5.6	2105	13	ADN60269	Adn60269 Human ZIP
38	168.6	5.6	2105	13	ADR83493	Adr83493 Human dea
39	168.6	5.6	2132	2	AAX34656	Aax34656 Human ZIP
40	168.6	5.6	2132	8	ACA90235	Aca90235 Deatch as
41	168.6	5.6	2132	12	ADO09424	Ado09424 Novel hum
42	168.6	5.6	2224	4	AAH16158	Aah16158 Human CDN
43	168.6	5.6	2224	5	AAH78068	Aah78068 Nucleotid
44	164	5.5	1514	10	ADB58463	Adb58463 Toxicity-
45	164	5.5	1514	10	ADB53052	Adb53052 Primary r

ALIGNMENTS

RESULT 1
AAH46904
ID AAH46904 standard; cDNA; 4936 BP.
XX
AC AAH46904;
XX
DT 25-SEP-2001 (first entry)
XX
DE cDNA encoding human protein kinase SGK145.

Protein kinase; enzyme; cytostatic; nootropic; neuroprotective; human;
antiparkinsonian; virucide; antibacterial; antifungal; antimigraine;
analgesic; hypotensive; hypertensive; immunosuppressive; anti-allergic;
antipruritic; antirheumatic; antithrombotic; ophthalmologic; anorectic;
osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;
vasotropic; antidiabetic; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FH 1. .4851
FT /*tag= a
FT

XX WO200155356-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US002337.

XX 25-JAN-2000; 2000US-0178078P.

XX 31-JAN-2000; 2000US-0179364P.

XX 17-FEB-2000; 2000US-0183173P.

XX 17-MAR-2000; 2000US-0190162P.

XX 29-MAR-2000; 2000US-0191404P.

XX 13-NOV-2000; 2000US-0247013P.

XX (SUGE-) SUGEN INC.

XX Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;

XX WPI; 2001-476202/51.

XX P-PSDB; AAB85504.

PT Kinase polypeptides useful for treating cancers, Alzheimer's disease,
PT viral infections, diabetes, obesity, organ transplant rejection and
XX rheumatoid arthritis.

XX Example 1; Page 204-205; 218pp; English.

CC The invention provides human protein kinases and protein kinase-like
CC enzymes and polynucleotides encoding the polypeptides. The kinase
CC polypeptides and their modulators are useful for treating a disease or
CC disorder such as cancer, immune-related diseases, cardiovascular disease,
CC brain or neuronal-associated disease and metabolic disorders, including
CC cancers of tissues, cancers of hematopoietic origin, diseases of the
CC central nervous system, diseases of the peripheral nervous system,
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
CC lateral sclerosis, viral infections, infections caused by prions,
CC bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction,
CC mood disorders, attention disorders, cognition disorders, hypotension,
CC hypertension, psychotic disorders, neurological disorders, dyskinesias,
CC metabolic disorders, and organ transplant rejection. They are also useful
CC for treating rhinitis, autoimmunity, atherosclerosis, psoriasis,
CC osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic
CC inflammatory bowel disease, rheumatoid arthritis, metabolic disorders
CC such as diabetes, obesity, cardiovascular diseases such as reperfusion
CC injury, coronary thrombosis, clotting disorders and atherosclerosis,
CC ocular diseases such as glaucoma, retinopathy and macular degeneration,
CC psychiatric and neurological disorders such as anxiety, schizophrenia,
CC dementia, manic depression, etc. The polynucleotides are useful in gene
CC therapy techniques to treat the above mentioned disorders. Sequences
CC AAH46891-46922 represent human protein kinases encoding cDNA molecules

XX Sequence 4936 BP; 923 A; 1695 C; 1524 G; 794 T; 0 U; 0 Other;

Query Match 100.0%; Score 3001; DB 4; Length 4936;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTAGGAGCAGAACTCGGGCCAGGATACAGGAGCGAGACATCTGCGCGCGCTGAG 60
DB 450 CCTACGAGCAGAACTCGGGCCAGGATACAGGAGCGAGACATCTGCGCGCGCTGAG 509
QY 61 CCACCCGCTGTACGCGGGCTGTGGACCACTTTGAGACCGCGAAGACCTCATCTCCAT 120
DB 510 CCACCCGCTGTACGCGGGCTGTGGACCACTTTGAGACCGCGAAGACCTCATCTCCAT 569
QY 121 CTTGAGCTGTCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAAGGCGTGTGAC 180
DB 570 CTTGAGCTGTCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAAGGCGTGTGAC 629
QY 181 GGAGCGCAGGTCAAGTCTACATCCAGAGCTGTGGAGGGGTGCACCTACCTGCACAG 240
DB 630 GGAGCGCAGGTCAAGTCTACATCCAGAGCTGTGGAGGGGTGCACCTACCTGCACAG 689
QY 241 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTAACATCTGATGTGTATCTGCGCG 300
DB 690 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTAACATCTGATGTGTATCTGCGCG 749
QY 301 GGAAGACATTAATAATCTGCGACTTTGGCTTTGGCCGAAACATCACCCGACAGAGTGTCA 360
DB 750 GGAAGACATTAATAATCTGCGACTTTGGCTTTGGCCGAAACATCACCCGACAGAGTGTCA 809
QY 361 GTTCAGCAGTACGGCTCCCTGAGTTGCTCTCCCGGAGATCATTCAGAGAACCTGT 420
DB 810 GTTCAGCAGTACGGCTCCCTGAGTTGCTCTCCCGGAGATCATTCAGAGAACCTGT 869
QY 421 GAGCGAGCTCCGACATTTGGGCGATGGGTGTCTCTTACCTACGACCTGACCTGCTC 480
DB 870 GAGCGAGCTCCGACATTTGGGCGATGGGTGTCTCTTACCTACGACCTGACCTGCTC 929
QY 481 ATCCCAATTTGCGGCGAGAGTGCACCGTGCACCCCTCTGAAACGCTCTGGAGGGGCGGT 540
DB 930 ATCCCAATTTGCGGCGAGAGTGCACCGTGCACCCCTCTGAAACGCTCTGGAGGGGCGGT 989
QY 541 GTATGGAGCAGCCCCATGCTGCCCACTTCAGGAGAGCGCCAAAGACTTTCATCAAGGC 600

DB 990 GTATGGAGCAGCCCCATGCTGCCCACTTCAGGAAAGAGCGCCAAAGACTTTCATCAAGGC 1049
QY 601 TAGGCTGCAGAGAGCCCCCTCAGGCCCCGGCCCTAGTGGCGGCCAGTGCCTCTCCCAACCCCTG 660
DB 1050 TAGGCTGCAGAGAGCCCCCTCAGGCCCCGGCCCTAGTGGCGGCCAGTGCCTCTCCCAACCCCTG 1109
QY 661 GTTCTGAAATTCATGCTCGGAGAGGCGCCACTTTCATCAACCAAGAGAGCTCAAGTT 720
DB 1110 GTTCTGAAATTCATGCTCGGAGAGGCGCCACTTTCATCAACCAAGAGAGCTCAAGTT 1169
QY 721 CTTCTGGCCGGAAGTCTGCGAGCGTTCCCTGATGAGCTACAGTCCCATCTCTGCTGAT 780
DB 1170 CTTCTGGCCGGAAGTCTGCGAGCGTTCCCTGATGAGCTACAGTCCCATCTCTGCTGAT 1229
QY 781 GCGCTCCATCCCTGAGCTGCTGGGGGCCCAACCCGACAGAGCCCTCCCTCGGCGTAGGCCG 840
DB 1230 GCGCTCCATCCCTGAGCTGCTGGGGGCCCAACCCGACAGAGCCCTCCCTCGGCGTAGGCCG 1289
QY 841 GCACCTCTGACGAGACACTGGTGGCTCCTCAAGTTCTCTCTCTCTCTCTCTCTGACAAAGCT 900
DB 1290 GCACCTCTGACGAGACACTGGTGGCTCCTCAAGTTCTCTCTCTCTCTCTCTCTGACAAAGCT 1349
QY 901 CGCCCAATTTGCCGGGCTAAGTCACTGCCACCTCCCGGTGACACACTCACACTGCT 960
DB 1350 CGCCCAATTTGCCGGGCTAAGTCACTGCCACCTCCCGGTGACACACTCACACTGCT 1409
QY 961 GCACCCCGGGGCTTCTGCGGGCCCTCGGCGAGCTCTCTGAGGAAGCCGAGGCGAGTGA 1020
DB 1410 GCACCCCGGGGCTTCTGCGGGCCCTCGGCGAGCTCTCTGAGGAAGCCGAGGCGAGTGA 1469
QY 1021 GCGCTCCACGAGGCGCCAGCTCGCTGCTATCTCCGAGGGTGCAGGGCGCACCGGCGC 1080
DB 1470 GCGCTCCACGAGGCGCCAGCTCGCTGCTATCTCCGAGGGTGCAGGGCGCACCGGCGC 1529
QY 1081 CAGGGCTGCTGCTCCCGGCGACAGCTCATCCGAGCTGTTTACCAACAGGCGGCTGA 1140
DB 1530 CAGGGCTGCTGCTCCCGGCGACAGCTCATCCGAGCTGTTTACCAACAGGCGGCTGA 1589
QY 1141 GAGCCCTGAGACAGCGGGCCCTGGCCCCGGGAGCAGGCGGCAACCGGCGCGGCGGCA 1200
DB 1590 GAGCCCTGAGACAGCGGGCCCTGGCCCCGGGAGCAGGCGGCAACCGGCGCGGCGGCA 1649
QY 1201 CTTGCTGAAGGGCGGCTACATTCGCGGGCGCTGCCAGGCTTCGCGAGCACTGTATGA 1260
DB 1650 CTTGCTGAAGGGCGGCTACATTCGCGGGCGCTGCCAGGCTTCGCGAGCACTGTATGA 1709
QY 1261 GCACCGGCTGCTGAGGAGGAGGCGCCGAGGAGGAGCAGGCCACCTCTGCGCAAGC 1320
DB 1710 GCACCGGCTGCTGAGGAGGAGGCGCCGAGGAGGAGCAGGCCACCTCTGCGCAAGC 1769
QY 1321 CCCTCATTCGAGACTGCCCTCCGGCTGCTGCTCTGGCAACCCACTTGGCCCCCTGGCCA 1380
DB 1770 CCCTCATTCGAGACTGCCCTCCGGCTGCTGCTCTGGCAACCCACTTGGCCCCCTGGCCA 1829
QY 1381 CAGCCACTCCCTGAAACATGACTCTCCGAGCACCCCCCGCCCTCTCTCGAGGCGCTCGG 1440
DB 1830 CAGCCACTCCCTGAAACATGACTCTCCGAGCACCCCCCGCCCTCTCTCGAGGCGCTCGG 1889
QY 1441 TGAGGACACAGGACTGCTTACGCCCCCTCCGGGGGGGGCCCCCTATCAGGACATCGGGCA 1500
DB 1890 TGAGGACACAGGACTGCTTACGCCCCCTCCGGGGGGGGCCCCCTATCAGGACATCGGGCA 1949
QY 1501 CCCTCAGGGCTCCAAAGCAGCTTCCATCCACTGCTGGCGCACCCAGGCACTGCTCAGCCAGA 1560
DB 1950 CCCTCAGGGCTCCAAAGCAGCTTCCATCCACTGCTGGCGCACCCAGGCACTGCTCAGCCAGA 2009
QY 1561 GAGGCATCCCGGACAGCCCTTGGGGGAGCAGCAGCCCTTTTCTGCCACCCAGAGAGGG 1620
DB 2010 GAGGCATCCCGGACAGCCCTTGGGGGAGCAGCAGCCCTTTTCTGCCACCCAGAGAGGG 2069
QY 1621 TTCTGCCCCCAGGAGGCTGACGCCCCCAGCCAGAGTTGCCCATGCCCCCTCTCTGCTC 1680

Db 2070 TTCTGCCCCCCAGGAGGGCTGCAGCCGCCACCCAGCAGTTGCCCCATGCCCCCTCTGGCTC 2129
Qy 1681 CTTCCCTTCAGGATCTTGAAGAAGGCCCCCTTAGTACCTCAAGCCCTTCTTGGGACA 1740
Db 2130 CTTCCCTTCAGGATCTTGAAGAAGGCCCCCTTAGTACCTCAAGCCCTTCTTGGGACA 2189
Qy 1741 GCCCCAGGACCCCTGCCTGCTGCAAGCAAGCCGCCCTTGGACTTGAAGTGGGGCC 1800
Db 2190 GCCCCAGGACCCCTGCCTGCTGCAAGCAAGCCGCCCTTGGACTTGAAGTGGGGCC 2249
Qy 1801 TGGAGACATCTCTCTTCTGAGAGGCCCCAAACCCCGGCCCTTGCAGTTTCCCCAGGGTTCAGC 1860
Db 2250 TGGAGACATCTCTCTTCTGAGAGGCCCCAAACCCCGGCCCTTGCAGTTTCCCCAGGGTTCAGC 2309
Qy 1861 CTCCAGGAGGATCTTCCCAAGTAGCTTCCCTCAGGGTGGGCTCTTCCCAAGTGGGGAC 1920
Db 2310 CTCCAGGAGGATCTTCCCAAGTAGCTTCCCTCAGGGTGGGCTCTTCCCAAGTGGGGAC 2369
Qy 1921 AGAGCCTGGCCCTCCCTGGATGGAGGGGCTGGACCCAGGAGCTGAGGATCTGCCGA 1980
Db 2370 AGAGCCTGGCCCTCCCTGGATGGAGGGGCTGGACCCAGGAGCTGAGGATCTGCCGA 2429
Qy 1981 CTCCACACCCACCTTGCAGCGCCCTCAGGAACAGCGCACCATGGCGAAGTTCTCCCTGGG 2040
Db 2430 CTCCACACCCACCTTGCAGCGCCCTCAGGAACAGCGCACCATGGCGAAGTTCTCCCTGGG 2489
Qy 2041 TGGTCGGGGGGCTACGAGCGGCTGGCTTGGCTATGGCACTTTGCTTGGTGAGATGC 2100
Db 2490 TGGTCGGGGGGCTACGAGCGGCTGGCTTGGCTATGGCACTTTGCTTGGTGAGATGC 2549
Qy 2101 AGGGGGATGCTGGGGCAGGGGCCATGCTGGCCAGGATAGCTGGGCTGTGCCAGTC 2160
Db 2550 AGGGGGATGCTGGGGCAGGGGCCATGCTGGCCAGGATAGCTGGGCTGTGCCAGTC 2609
Qy 2161 GGAGGAGGAGGAGGAGGAGGCGCAGCGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2220
Db 2610 GGAGGAGGAGGAGGAGGAGGCGCAGCGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2669
Qy 2221 CAGGGCTGAGAGGCCACTGCCCCAGGTCAGTGCAGGCGCTGTGCTGAGGTTCGAGGCG 2280
Db 2670 CAGGGCTGAGAGGCCACTGCCCCAGGTCAGTGCAGGCGCTGTGCTGAGGTTCGAGGCG 2729
Qy 2281 TCCACACGAGGATCTCCAGAGCCACCCCATGGAGGACATCGGGCAGGTCTCCCTGGT 2340
Db 2730 TCCACACGAGGATCTCCAGAGCCACCCCATGGAGGACATCGGGCAGGTCTCCCTGGT 2789
Qy 2341 GCAGATCCGGACCTGTGAGTGATGCGAGGCGCGCACCAATATCCCTGGACATTTTC 2400
Db 2790 GCAGATCCGGACCTGTGAGTGATGCGAGGCGCGCACCAATATCCCTGGACATTTTC 2849
Qy 2401 CGAGGTGGAACCCCGCTACCTCAACCTCTCAGACCTGTGATATCAAGTACCTCCCAT 2460
Db 2850 CGAGGTGGAACCCCGCTACCTCAACCTCTCAGACCTGTGATATCAAGTACCTCCCAT 2909
Qy 2461 CGAGTTTATGATCTTCAGGAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCAT 2520
Db 2910 CGAGTTTATGATCTTCAGGAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCAT 2969
Qy 2521 GGCTGAGGAGGAGTGGCCGAGTTCCCGGAGCCACAGTGGCCCTGGCCAGGTGAACCTGGG 2580
Db 2970 GGCTGAGGAGGAGTGGCCGAGTTCCCGGAGCCACAGTGGCCCTGGCCAGGTGAACCTGGG 3029
Qy 2581 CCCCCACGAGGCTGGAGATCACAGAGGAGTGCAGAGATGTGACGCGCTGTCTGGCAGA 2640
Db 3030 CCCCCACGAGGCTGGAGATCACAGAGGAGTGCAGAGATGTGACGCGCTGTCTGGCAGA 3089
Qy 2641 GGCTGCGGTGGCAGGAGCGCAAGTGGTCTCCGCTCAGCAGCGCTTCTTCCACTTCCC 2700
Db 3090 GGCTGCGGTGGCAGGAGCGCAAGTGGTCTCCGCTCAGCAGCGCTTCTTCCACTTCCC 3149
Qy 2701 TGGAGGACCTGCGCGCTGGATGAGCTGCGAGAGCTGGGGCTGCGTGAGAGAGTGAAGGC 2760
Db 3150 TGGAGGACCTGCGCGCTGGATGAGCTGCGAGAGCTGGGGCTGCGTGAGAGAGTGAAGGC 3209

RESULT 2

ABX11641

ID ABX11641 standard; cDNA; 8106 BP.

XX AC ABX11641;

XX DT 09-MAY-2003 (first entry)

XX DE Human serine/threonine or protein kinase 59079, cDNA.

XX KW Human; ss; gene; serine/threonine kinase; protein kinase; 59079;
cardiovascular disease; heart failure; myocardial infarction;
XX KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
XX KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
XX KW haemolytic anaemia; cellular proliferative disorder; cancer;
XX KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
XX KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
multiple sclerosis.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT 5'UTR 1..71

FT CDS /*tag= a

FT /*tag= b

FT /*tag= c

FT /product= "Kinase 59079"

FT /note= "This CDS is specifically claimed in claim 2"

FT 3'UTR 7965..8106

FT /*tag= c

PN US2002168742-A1.

XX PD 14-NOV-2002.

XX PF 15-FEB-2002; 2002US-00077130.

XX PR 15-FEB-2001; 2001US-0269201P.

XX XX (MILL-) MILLENNIUM PHARM INC.

XX XX Kapeller-Libermann R, Acton SL;

XX XX WPI; 2003-298729/29.

XX XX P-PSDB; ABG76186.

XX PT Novel isolated human protein kinase, designated 59079 or 12599
polypeptide, useful as diagnostic and therapeutic agents for preventing
XX PT cardiovascular diseases, proliferative disorders, and protein kinase
disorders.

Db 5177 TTCTGCCCCCAGAGGGCTGAGCCCCCACCAGCAGTTGCCCATGCCCTCTCGGCTC 5236
Qy 1681 CTTCCCTCCAGGATCTTCAAGAGGCCCCCTTAGTACCTCAAGCCCCCTTCTTGGGACA 1740
Db 5237 CTTCCCTCCAGGATCTTCAAGAGGCCCCCTTAGTACCTCAAGCCCCCTTCTTGGGACA 5296
Qy 1741 GCCCAGGACCCCTGCCCCCTGCCAAGAGAGCCCCCCATTTGACTCTAAGATGGGGCC 1800
Db 5297 GCCCAGGACCCCTGCCCCCTGCCAAGAGAGCCCCCCATTTGACTCTAAGATGGGGCC 5356
Qy 1801 TGGAGACATCTCTCTCTCTGGAGGCCCCAAACCCGGCCCCCTGCACTGCCAGAGGTGACG 1860
Db 5357 TGGAGACATCTCTCTCTGGAGGCCCCAAACCCGGCCCCCTGCACTGCCAGAGGTGACG 5416
Qy 1861 CTCCAGGCGAGCTCTTCCCAAGTGAAGTCTCCCTCAGAGGTGGCTCTCCCAAGGTGGGAC 1920
Db 5417 CTCCAGGCGAGCTCTTCCCAAGTGAAGTCTCCCTCAGAGGTGGCTCTCCCAAGGTGGGAC 5476
Qy 1921 AGAGCCTGGCCCCCTCCCTGGATGCGGAGGGCTGGACCCAGGAGGTGAGGATCTGTCCGA 1980
Db 5477 AGAGCCTGGCCCCCTCCCTGGATGCGGAGGGCTGGACCCAGGAGGTGAGGATCTGTCCGA 5536
Qy 1981 CTCACACCCACCTTGACGCGGCTCAGAAACAGGCGACCATGCGCAAGTTCTCCCTGGG 2040
Db 5537 CTCACACCCACCTTGACGCGGCTCAGAAACAGGCGACCATGCGCAAGTTCTCCCTGGG 5596
Qy 2041 TGGTCGCGGGGGCTACGAGCGGTGGCTGGCTATGGCACCTTTGCTTTGGTGGAGATGC 2100
Db 5597 TGGTCGCGGGGGCTACGAGCGGTGGCTGGCTATGGCACCTTTGCTTTGGTGGAGATGC 5656
Qy 2101 AGGGGGCATGTGGGGCAGGGGCCCATGTGGCCAGGATAGCTGGGCTGTGTCCAGTC 2160
Db 5657 AGGGGGCATGTGGGGCAGGGGCCCATGTGGCCAGGATAGCTGGGCTGTGTCCAGTC 5716
Qy 2161 GGAGGAGGAGCAGGAGGAGGCGCAGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 2220
Db 5717 GGAGGAGGAGCAGGAGGAGGCGCAGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 5776
Qy 2221 CAGGCTGAGAGGCCACTGCCCCAGGTGAGTGCAGAGGCTGTGCTGAGGTGCGGAGGCG 2280
Db 5777 CAGGCTGAGAGGCCACTGCCCCAGGTGAGTGCAGAGGCTGTGCTGAGGTGCGGAGGCG 5836
Qy 2281 TCCACACGAGGCTCTCAGAGCCACCCCATGGAGAGACATCGGGCAGGTCTCCCTGGT 2340
Db 5837 TCCACACGAGGCTCTCAGAGCCACCCCATGGAGAGACATCGGGCAGGTCTCCCTGGT 5896
Qy 2341 GCAGATCCGGACCTGTTCAGGTGATGCGGAGCGCGCACACATATCCCTGGACATTC 2400
Db 5897 GCAGATCCGGACCTGTTCAGGTGATGCGGAGCGCGCACACATATCCCTGGACATTC 5956
Qy 2401 CGAGGTGGACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db 5957 CGAGGTGGACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 6016
Qy 2461 CGAGTTTATGATCTTCAGAAAGTCCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 2520
Db 6017 CGAGTTTATGATCTTCAGAAAGTCCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 6076
Qy 2521 GGCTGAGGAGGCTGGCGGAGTTCCCGAGCCACACGTGGCCCTGGCCAGGTGAAGTGGG 2580
Db 6077 GGCTGAGGAGGCTGGCGGAGTTCCCGAGCCACACGTGGCCCTGGCCAGGTGAAGTGGG 6136
Qy 2581 CCCCCACGAGCCTGGAGATCACAGAGGAGTCAGAGGATGTGGACGCGCTGTGGCAGA 2640
Db 6137 CCCCCACGAGCCTGGAGATCACAGAGGAGTCAGAGGATGTGGACGCGCTGTGGCAGA 6196
Qy 2641 GGCTGCTGGGCGAGGAAGCGCAAGTGTCTTCCGCTCAGCAGACCTCTTCCACTTCCC 2700
Db 6197 GGCTGCTGGGCGAGGAAGCGCAAGTGTCTTCCGCTCAGCAGACCTCTTCCACTTCCC 6256
Qy 2701 TGGGAGGACCTGCGGCTGGATGAGCCTGACAGGTGGGGCTGCGTGAAGAGTGAAGGC 2760

Db 6257 TGGGAGGACCTGCGGCTGGATGAGCCTGCAGAGCTGGGGCTGCGTGAAGAGTGAAGGC 6316
Qy 2761 CTCGCTGGAGCACATCTCCCGATCTCGGATCTGAAAGGCGAGCCGGAAGGTCTGGAGAGGAGGG 2820
Db 6317 CTCGCTGGAGCACATCTCCCGATCTCGGATCTGAAAGGCGAGCCGGAAGGTCTGGAGAGGAGGG 6376
Qy 2821 GCCCCCCCAGGAGAGAGCCAGCCTTGTCTTCCCTCCGGCTCTCAGGTCTGAAGAGCTGGGA 2880
Db 6377 GCCCCCCCAGGAGAGAGCCAGCCTTGTCTTCCCTCCGGCTCTCAGGTCTGAAGAGCTGGGA 6436
Qy 2881 CCGAGCGCCGACATCTCTTAAGGAGCTCTCAGATGAGACTGTGGTCTCTGGGCGAGTCAGT 2940
Db 6437 CCGAGCGCCGACATCTCTTAAGGAGCTCTCAGATGAGACTGTGGTCTCTGGGCGAGTCAGT 6496
Qy 2941 GACACTGGCTTCAGAGGTTCAGCCCGAGCTGCCCCAGGCCACTTCGAGCAAGAGCGG 3000
Db 6497 GACACTGGCTTCAGAGGTTCAGCCCGAGCTGCCCCAGGCCACTTCGAGCAAGAGCGG 6556
Qy 3001 A 3001
Db 6557 A 6557
RESULT 3
ABX11642
ID ABX11642 standard; cDNA; 24120 BP.
XX
AC ABX11642;
XX
DT 09-MAY-2003 (first entry)
XX
DE Human serine/threonine or protein kinase 12599, cDNA.
XX
KW Human; ss; gene; serine/threonine kinase; protein kinase; 12599;
cardiovascular disease; heart failure; myocardial infarction;
blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
blood platelet disorder; thrombocytopenia; leukaemia; Hodgkin's disease;
haemolytic anaemia; cellular proliferative disorder; cancer;
protein kinase disorder; autoimmune disorder; diabetes mellitus;
psoriasis; inflammatory bowel disease; rheumatoid arthritis;
multiple sclerosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..71
FT CDS /*tag= a
FT 72..23978
FT /*tag= b
FT /product= "Kinase 12599"
FT /note= "This CDS is specifically claimed in claim 2"
FT 3'UTR 23979..24120
FT /*tag= c
XX
PN US2002168742-A1.
XX
PD 14-NOV-2002.
XX
PF 15-FEB-2002; 2002US-00077130.
XX
PR 15-FEB-2001; 2001US-0269201P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Kapeller-Libermann R, Acton SL;
XX
XX WPI; 2003-298729/29.
DR P-PSDB; ABG76187.
XX
PT Novel isolated human protein kinase, designated 59079 or 12599
polypeptide, useful as diagnostic and therapeutic agents for preventing
cardiovascular diseases, proliferative disorders, and protein kinase
disorders.

XX

Claim 2; Page 58-84; 119pp; English.

The invention relates to an isolated human serine/threonine or protein kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule comprising at least 85% identity to the nucleic acids appearing as ABX11641 and ABX11642 or their complement, a naturally occurring variant of the kinases or their fragments. Also included are a non-human host cell containing the nucleic acids, an antibody specific for the proteins, identifying a compound which binds to the kinase (by contacting the kinase or a cell expressing the kinase with a test compound and determining whether the kinase binds to the test compound) and modulating the activity of kinase using the identified compound. The kinases and their encoding nucleic acids are useful as diagnostic and therapeutic agents for preventing a disease or condition associated with an aberrant or unwanted 59079 or 12599 activity in a subject, including cardiovascular diseases such as heart failure, and myocardial infarction; disorders involving blood vessels such as atherosclerosis, and Kaposi's sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia, Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders such as cancer; and protein kinase disorders such as autoimmune disorders, diabetes mellitus, psoriasis, inflammatory bowel disease, rheumatoid arthritis, and multiple sclerosis (many examples of diseases and disorders are included in the specification). The kinases, their encoding nucleic acids and antibodies are useful in screening assays, detection assays (e.g. forensic biology), and predictive medicine (e.g. diagnostic assays, prognostic assays, and monitoring clinical trials and pharmacogenomics). The kinases and their encoding nucleic acids are useful as query sequences to perform a search against public databases to identify other family members or related sequences. The present sequence encodes the kinase 12599

Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;

Query Match	100.0%;	Score 3001;	DB 8;	Length 24120;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3001;	Conservative 0;	Mismatches 0;	Indels 0;	

Qy	1	CCTACGAGCAGAACTCGGGCCCGAGGATACAGGAGCGAGACATCTCTGGCCGCGCTGAG	60
Db	19571	CCTACGAGCAGAACTCGGGCCCGAGGATACAGGAGCGAGACATCTCTGGCCGCGCTGAG	19630
Qy	61	CCACCCGCTGCTCA CGGGGCTGTGGAC CAGTTTGAGACCCGCAAGACCCCTCATCTCAT	120
Db	19631	CCACCCGCTGCTCA CGGGGCTGTGGAC CAGTTTGAGACCCGCAAGACCCCTCATCTCAT	19690
Qy	121	CCTGGAGCTGTGCTCATCCGAGAGCTGTGGACCGCCTGTACAGAAAGGCGTGGTGAC	180
Db	19691	CCTGGAGCTGTGCTCATCCGAGAGCTGTGGACCGCCTGTACAGAAAGGCGTGGTGAC	19750
Qy	181	GGAGGCCGAGGTC AAGGTC TACATCCAGCAGCTGGTGGAGGSGCTGCATCTCTGCACAG	240
Db	19751	GGAGGCCGAGGTC AAGGTC TACATCCAGCAGCTGGTGGAGGSGCTGCATCTCTGCACAG	19810
Qy	241	CNATGGGTTCTCCA CTGGACATAAAGCCCTTAACATCTGTATGGTGCACTCTGCCCG	300
Db	19811	CNATGGGTTCTCCA CTGGACATAAAGCCCTTAACATCTGTATGGTGCACTCTGCCCG	19870
Qy	301	GGAAGACATTTAAATCTGCGACCTTTGGCTTCCCGAGACATCACCCAGCAGAGCTGCA	360
Db	19871	GGAAGACATTTAAATCTGCGACCTTTGGCTTCCCGAGACATCACCCAGCAGAGCTGCA	19930
Qy	361	GTTTCAGCCAGTAGCGGCTCCCTGAGTTCGTCCTCCCGAGATCATCCAGCAGAACCTGTG	420
Db	19931	GTTTCAGCCAGTAGCGGCTCCCTGAGTTCGTCCTCCCGAGATCATCCAGCAGAACCTGTG	19990
Qy	421	GAGCGAAGCTCCGACATTTGGGCCATGGGTGTCATCTCCTACCTCAGCTGACCTGCTC	480
Db	19991	GAGCGAAGCTCCGACATTTGGGCCATGGGTGTCATCTCCTACCTCAGCTGACCTGCTC	20050
Qy	481	ATCCCCCATTTGCCGGCGAGTGACCGTGCCACCTCTCTGAACTGTCGGAGGGCGCGGT	540
Db	20051	ATCCCCCATTTGCCGGCGAGTGACCGTGCCACCTCTCTGAACTGTCGGAGGGCGCGGT	20110

Qy 1621 TTCTGCCCCCCAGAGGGCTGCAGCCGCCCCAGCAGAGTTGCCCCATGCCCTCTGGTCTC 1680
Db 21191 TTCTGCCCCCCAGAGGGCTGCAGCCGCCCCAGCAGAGTTGCCCCATGCCCTCTGGTCTC 21250
Qy 1681 CTTCCCTCCAGATCTTTGCAAGAGGGCCCCCTTAGTACCCCTCAAGCCCTTCTTTGGGACA 1740
Db 21251 CTTCCCTCCAGATCTTTGCAAGAGGGCCCCCTTAGTACCCCTCAAGCCCTTCTTTGGGACA 21310
Qy 1741 GCCCCAGGCAACCCCTGCTCCCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 21311 GCCCCAGGCAACCCCTGCTCCCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21370
Qy 1801 TGGAGATCTCTCTCTCTGGAGGAGCAAAACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 21371 TGGAGATCTCTCTCTCTGGAGGAGCAAAACCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 21430
Qy 1861 CTCCAGGAGAGCTCTTCCCAAGTAGCTCTCCCTCAGAGGTGGGCTCTCCAGAGTGGGCAC 1920
Db 21431 CTCCAGGAGAGCTCTTCCCAAGTAGCTCTCCCTCAGAGGTGGGCTCTCCAGAGTGGGCAC 21490
Qy 1921 AGAGCTTGGCCCTCTCTGGAG 1980
Db 21491 AGAGCTTGGCCCTCTCTGGAG 21550
Qy 1981 CTCCACACCCACCTTGAGCGGCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db 21551 CTCCACACCCACCTTGAGCGGCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21610
Qy 2041 TGGTCCGGGGGCTTACGAGCGCTGGCTGGCTATGGGACCTTTGGCTTTGGTGAGATGC 2100
Db 21611 TGGTCCGGGGGCTTACGAGCGCTGGCTGGCTATGGGACCTTTGGCTTTGGTGAGATGC 21670
Qy 2101 AGGGGGGATGCTGGGGCAGAGGGCCCATGTGGGCGAGATAGCCTTGGGCTGTGTCCAGATC 2160
Db 21671 AGGGGGGATGCTGGGGCAGAGGGCCCATGTGGGCGAGATAGCCTTGGGCTGTGTCCAGATC 21730
Qy 2161 GGAGGAG 2220
Db 21731 GGAGGAG 21790
Qy 2221 CAGGGCTGAGAGCCACTTGCCCGCAGAGTCAGTGCAGAGGCTGTGCTGAGGTGCGCAGGGC 2280
Db 21791 CAGGGCTGAGAGCCACTTGCCCGCAGAGTCAGTGCAGAGGCTGTGCTGAGGTGCGCAGGGC 21850
Qy 2281 TCCACACGAGAGCTCTCAGAGCCACCCCATGGAGAGACATCGGGCAGGTCTCCCTGGT 2340
Db 21851 TCCACACGAGAGCTCTCAGAGCCACCCCATGGAGAGACATCGGGCAGGTCTCCCTGGT 21910
Qy 2341 GCAGATCGGAGACCTGTGAGTGATGCGGAGGCGGCGGAGACATATCCCTGGACATTTTC 2400
Db 21911 GCAGATCGGAGACCTGTGAGTGATGCGGAGGCGGCGGAGACATATCCCTGGACATTTTC 21970
Qy 2401 CGAGGTGGAGCCCGCTTACCTCAACCTCTCAGACTGTGACGATATCAAGTACCTCCCATTT 2460
Db 21971 CGAGGTGGAGCCCGCTTACCTCAACCTCTCAGACTGTGACGATATCAAGTACCTCCCATTT 22030
Qy 2461 CGAGTTTATGATTTTCAAGAAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCCAT 2520
Db 22031 CGAGTTTATGATTTTCAAGAAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCCAT 22090
Qy 2521 GGCTGAGGAGAGCTGGCGAGTTTCCCGAGGCCACAGTGGCCCTGGCCAGGTGAATGGG 2580
Db 22091 GGCTGAGGAGAGCTGGCGAGTTTCCCGAGGCCACAGTGGCCCTGGCCAGGTGAATGGG 22150
Qy 2581 CCCCCAGCGAGCTGGAGATCACAGAGAGAGTCAAGAGATGTGGACGGCTGTCTGGCAGA 2640
Db 22151 CCCCCAGCGAGCTGGAGATCACAGAGAGAGTCAAGAGATGTGGACGGCTGTCTGGCAGA 22210
Qy 2641 GGCTGCCGTGGGAGAGCGCAAGTGTCTCTCCCGTACGAGAGCTCTTCCACTTCCC 2700
Db 22211 GGCTGCCGTGGGAGAGCGCAAGTGTCTCTCCCGTACGAGAGCTCTTCCACTTCCC 22270
Qy 2701 TGGAGGACCTCGCGCTGGATGAGCCTGCAGAGCTGGGGCTGGTGAGAGAGTGAAGGC 2760

Db 22271 TGGAGGACCTCGCGCTGGATGAGCCTGCAGACTGGGGCTGGTGAGAGTGAAGGC 22330
Qy 2761 CTCCGTGGAGCACATCTCCCGGATCTTGAAGGGCAGGCCGGAAGGTCTTGGAGAAGGAGG 2820
Db 22331 CTCCGTGGAGCACATCTCCCGGATCTTGAAGGGCAGGCCGGAAGGTCTTGGAGAAGGAGG 22390
Qy 2821 GCCCCCAGGAG 2880
Db 22391 GCCCCCAGGAG 22450
Qy 2881 CCAGAGCCCGACATCTCTAAGGAGAGTCTCAGATGAGACTGTGTCTCTGGGCCAGTCAGT 2940
Db 22451 CCAGAGCCCGACATCTCTAAGGAGAGTCTCAGATGAGACTGTGTCTCTGGGCCAGTCAGT 22510
Qy 2941 GACACTGCGCTGCCAGAGTGTGAGCCAGCCAGCTGCCAGGCCACCTGGAGCAAGAGCGG 3000
Db 22511 GACACTGCGCTGCCAGAGTGTGAGCCAGCCAGCTGCCAGGCCACCTGGAGCAAGAGCGG 22570
Qy 3001 A 3001
Db 22571 A 22571
RESULT 4
AAC62286
ID AAC62286 standard; cDNA; 5007 BP.
XX AAC62286;
XX 19-MAR-2001 (first entry)
DT
XX
DE cDNA encoding a full length human signal transduction polypeptide.
XX
KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
KW congestive heart failure; dilated congestive cardiomyopathy;
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW mitral valve disease; aortic valve disease; tricuspid valve disease;
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW atherosclerosis; cardiac tumour; microbial infection; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 97..4929
FT FT /*tag= a
FT FT /product= "signal transduction polypeptide H19G5"
XX
XX WO200063381-A1.
XX
XX 26-OCT-2000.
XX
XX 11-APR-2000; 2000WO-US009488.
XX
XX 16-APR-1999; 99US-0129553P.
XX
XX (SCIO-) SCIOS INC.
XX PA
XX Zeng W, Stanton L, Kong H;
XX
XX WPI; 2001-007013/01.
XX P-PSDB; AAB30568.
XX
XX Novel h19G5 polypeptides capable of regulating signal transduction and
XX exhibiting kinase activity useful for identifying antibodies to treat
XX cardiac diseases, and additional mediators of signal transduction.
PS Claim 4; Page 59-61; 81pp; English.
XX
XX The present sequence encodes a human protein with putative function in
XX signal transduction. The polypeptide is designated H19G5. The protein is
XX capable of regulating signal transduction and exhibits kinase activity.
XX The H19G5 transcript is expressed in the heart. H19G5 polypeptides and

CC polynucleotides are useful for preventing or treating a cardiac disease,
CC such as congestive heart failure, dilated congestive cardiomyopathy,
CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve
CC disease, aortic valve disease or tricuspid valve disease, angina
CC pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial
CC or rheovascular hypertension, arteriosclerosis, atherosclerosis and
CC cardiac tumours in humans. The polypeptide is also useful for detecting
CC the expression of a protein capable of regulating signal transduction or
CC the expression of a protein capable of acting as a donor or acceptor
CC molecule of a phosphate group. The monoclonal antibodies can be used as
CC probes for detecting discrete antigens expressed by tissue or cell
CC samples, and therefore used in humans for localization and monitoring of
CC microbial infection
XX
SQ Sequence 5007 BP; 946 A; 1722 C; 1541 G; 798 T; 0 U; 0 Other;

Query Match 99.9%; Score 2999.4; DB 5; Length 5007;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTACGGAGCAGAACTCGGGCCCGAGGCATACAGGAGCGAGACATCTCGGCGCGCTGAG 60
DB |||||
522 CCTACGGAGCAGAACTCGGGCCCGAGGCATACAGGAGCGAGACATCTCGGCGCGCTGAG 581
QY 61 CCACCCGCTGTACCGGGCTGTGGACGAGTTTGAGACCCGCAAGACCCCTCATCTCAT 120
DB |||||
582 CCACCCGCTGTGTACCGGGCTGTGGACGAGTTTGAGACCCGCAAGACCCCTCATCTCAT 641
QY 121 CTTGAGAGTGTCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAGGCGTGTGTAC 180
DB |||||
642 CTTGAGAGTGTCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAGGCGTGTGTAC 701
QY 181 GGAGCCGAGGTCAAGGTCTACATCCAGACGTGTGTGAGGGGTGTACACTACCTGACAG 240
DB |||||
702 GGAGCCGAGGTCAAGGTCTACATCCAGACGTGTGTGAGGGGTGTACACTACCTGACAG 761
QY 241 CCATGGCGTTCTCACCTGGACATTAAGCCCTCTAAACATCTGTATGTGATCTGCCCC 300
DB |||||
762 CCATGGCGTTCTCACCTGGACATTAAGCCCTCTAAACATCTGTATGTGATCTGCCCC 821
QY 301 GGAAGACATTAATAATCTGCGACTTTGGCTTTGGCCAGAACATACCCAGCAGAGTGCA 360
DB |||||
822 GGAAGACATTAATAATCTGCGACTTTGGCTTTGGCCAGAACATACCCAGCAGAGTGCA 881
QY 361 GTTCAGCCAGTACGGCTCCCTGAGTTCGTCTCCCGAGATCATCCAGCAGAACCTGT 420
DB |||||
882 GTTCAGCCAGTACGGCTCCCTGAGTTCGTCTCCCGAGATCATCCAGCAGAACCTGT 941
QY 421 GAGGGAAGCTTCCGACATTTGGGCCATGGGTGTCTCTCTACTCTAGCCTGACCTGCTC 480
DB |||||
942 GAGGGAAGCTTCCGACATTTGGGCCATGGGTGTCTCTCTACTCTAGCCTGACCTGCTC 1001
QY 481 ATCCCCATTTGCCGGCGAGAGTGACCCGTGTCACCTCTCTGAAAGTCTCTGGAGGGCGGT 540
DB |||||
1002 ATCCCCATTTGCCGGCGAGAGTGACCCGTGTCACCTCTCTGAAAGTCTCTGGAGGGCGGT 1061
QY 541 GTCATGAGCAGCCCCATGCTGCTGCCCATCTCAGGGAAGCGCCAAAGATTTCATCAAGGC 600
DB |||||
1062 GTCATGAGCAGCCCCATGCTGCTGCCCATCTCAGGGAAGCGCCAAAGATTTCATCAAGGC 1121
QY 601 TAGCTGTCAGAGAGCCCTCAGGCCCGGCTAGTGGGGCCAGTGTCTCTCCACCCCTG 660
DB |||||
1122 TAGCTGTCAGAGAGCCCTCAGGCCCGGCTAGTGGGGCCAGTGTCTCTCCACCCCTG 1181
QY 661 GTTCTCGAATTCATGCTGCGAGGAGGCCCATCTTCATCAACACCAAGCAGCTCAAGTT 720
DB |||||
1182 GTTCTCGAATTCATGCTGCGAGGAGGCCCATCTTCATCAACACCAAGCAGCTCAAGTT 1241
QY 721 CTTCTTGGCCCGAAGTGTGCTGGCAGCGTTCCCTGTATGAGCTCAAGTCCATCTGTGTAT 780
DB |||||
1242 CTTCTTGGCCCGAAGTGTGCTGGCAGCGTTCCCTGTATGAGCTCAAGTCCATCTGTGTAT 1301
QY 781 GGGCTCCATCTCTGAGTGTGCTGGGGGCCACCCGACAGCCCTCTCTCGGGGTAGCCCG 840

DB |||||
1302 GCGTCCATCCCTGAGTGTCTGGGGGCCACCCGACAGCCCTCTCTCGGCGTAGCCCG 1361
QY 841 GCACCTCTGCAGGAGACACTGTGTGCTCTCCAGTTCTCTCTCTCTCTGACACAGAGCT 900
DB |||||
1362 GCACCTCTGCAGGAGACACTGTGTGCTCTCCAGTTCTCTCTCTCTCTGACACAGAGCT 1421
QY 901 CGCCCATTTGCTCGGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACCACCTGCT 960
DB |||||
1422 CGCCCATTTGCTCGGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACCACCTGCT 1481
QY 961 GCACCCCGGGGCTTCTGCGGGCCCTCGGCGAGCTCTCTGAGGAAGCCGAGGCGAGTGA 1020
DB |||||
1482 GCACCCCGGGGCTTCTGCGGGCCCTCGGCGAGCTCTCTGAGGAAGCCGAGGCGAGTGA 1541
QY 1021 GCGTCCACCGAGGCCCTCAGCTCCGCTGCTATCTCCCGAGGGTGCAGGGCCACCGGGCCG 1080
DB |||||
1542 GCGTCCACCGAGGCCCTCAGCTCCGCTGCTATCTCCCGAGGGTGCAGGGCCACCGGGCCG 1601
QY 1081 CAGGGCTGCTGCTCCCGGACAGCGTCACTCCGAGCTCTTCTACACACGAGCGGGTGA 1140
DB |||||
1602 CAGGGCTGCTGCTCCCGGACAGCGTCACTCCGAGCTCTTCTACACACGAGCGGGTGA 1661
QY 1141 GAGCCTCTGAGCAGCGGGCCCTGCGGCGAGCAGCGCACCCGCGCGGGCGGGCA 1200
DB |||||
1662 GAGCCTCTGAGCAGCGGGCCCTGCGGCGAGCAGCGCACCCGCGCGGGCGGGCA 1721
QY 1201 CTTGCTGAAGGGGGGTATCATTTGCGGGGCGCTGTCAGGCGCTGCGGAGCCTATGATGA 1260
DB |||||
1722 CTTGCTGAAGGGGGGTATCATTTGCGGGGCGCTGTCAGGCGCTGCGGAGCCTATGATGA 1781
QY 1261 GCACCCGCTGCTGAGGAGGAGGCGCGCAGGAGGAGCAGGCGCACCTCTCTGGCCAAAGC 1320
DB |||||
1782 GCACCCGCTGCTGAGGAGGAGGCGCGCAGGAGGAGCAGGCGCACCTCTCTGGCCAAAGC 1841
QY 1321 CCGCTCATTTGAGACTGCGCTCTCGGCTGCTGCTCTGCGACCCACTTGGCCCTCTGGCCA 1380
DB |||||
1842 CCGCTCATTTGAGACTGCGCTCTCGGCTGCTGCTCTGCGACCCACTTGGCCCTCTGGCCA 1901
QY 1381 CAGCCACTCTCTGGAACATGACTCTCGAGCACCCCGCCCTCTCTCGAGGCGCTGCGG 1440
DB |||||
1902 CAGCCACTCTCTGGAACATGACTCTCGAGCACCCCGCCCTCTCTCGAGGCGCTGCGG 1961
QY 1441 TGAGGACACAGCAGTCTGCTTTCAGCCCTCTCGGGGGGGGGGGGGGGGGGGGGGGGG 1500
DB |||||
1962 TGAGGACACAGCAGTCTGCTTTCAGCCCTCTCGGGGGGGGGGGGGGGGGGGGGGGGG 2021
QY 1501 CCGCTCAGGGCTCCAAAGCAGCTTCCATCCACTGCTGGTGGCCACCCAGGCACTGTCTCAGCCAGA 1560
DB |||||
2022 CCGCTCAGGGCTCCAAAGCAGCTTCCATCCACTGCTGGTGGCCACCCAGGCACTGTCTCAGCCAGA 2081
QY 1561 GAGGCCATCCCGGAGACGCGCTTGGGGGCGAGCAGCCCTCTTCTGCGACCCCGCAAGCAGGG 1620
DB |||||
2082 GAGGCCATCCCGGAGACGCGCTTGGGGGCGAGCAGCCCTCTTCTGCGACCCCGCAAGCAGGG 2141
QY 1621 TTCTGCCCCCGCAGGAGGGCTGCGAGCCCCCAGCCAGCAGTGTGCCCATGCGCTCTCTGGGTC 1680
DB |||||
2142 TTCTGCCCCCGCAGGAGGGCTGCGAGCCCCCAGCCAGCAGTGTGCCCATGCGCTCTCTGGGTC 2201
QY 1681 CTTTCCCTCAGGATTTGCAAGAGGCGCCCTTATGATACCTCAAGCCCTCTTCTTGGGACA 1740
DB |||||
2202 CTTTCCCTCAGGATTTGCAAGAGGCGCCCTTATGATACCTCAAGCCCTCTTCTTGGGACA 2261
QY 1741 GCGCCAGGACACCCCTGCGCTCCAAAGCAGGAGGCGCCCTTATGATACCTCAAGCCCTCTTCTTGGGAG 1800
DB |||||
2262 GCGCCAGGACACCCCTGCGCTCCAAAGCAGGAGGCGCCCTTATGATACCTCAAGCCCTCTTCTTGGGAG 2321
QY 1801 TGAGAGACATCTCTCTTCTGAGGAGGCCAAAGACCCGGGGGGGGGGGGGGGGGGGGGGGG 1860
DB |||||
2322 TGAGAGACATCTCTCTTCTGAGGAGGCCAAAGACCCGGGGGGGGGGGGGGGGGGGGGGGG 2381
QY 1861 CTTCCGAGGCGAGCTTTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCTCCAGGTTGGGAC 1920

Db 2382 CTCCAGGCGAGCTCTTCCCAAGTGAGCTCCCTCAGGGTGGGCTCCTCCAGGTGGGCAC 2441
Qy 1921 AGAGCTTGCGCCCTCCCTTGATGCGAGGGCTGACCCAGAGGCTGAGGATCTGTGCCGA 1980
Db 2442 AGAGCTTGCGCCCTCCCTTGATGCGAGGGCTGACCCAGAGGCTGAGGATCTGTGCCGA 2501
Qy 1981 CTCCACACCCACCTTGAGCGGCGCTCAGGACAGCGGCATGCGAAGTTCTCCCTGGG 2040
Db 2502 CTCCACACCCACCTTGAGCGGCGCTCAGGACAGCGGCATGCGAAGTTCTCCCTGGG 2561
Qy 2041 TGGTTCGGGGGGCTACGACGCGTGGCTGATGCGACCTTTGCTTTGGTGAGATGC 2100
Db 2562 TGGTTCGGGGGGCTACGACGCGTGGCTGATGCGACCTTTGCTTTGGTGAGATGC 2621
Qy 2101 AGGGGGATGCTGGGGGAGGGGCCATGTGGGCGAGATAGCCTGGGCTGTGTCCAGTC 2160
Db 2622 AGGGGGATGCTGGGGGAGGGGCCATGTGGGCGAGATAGCCTGGGCTGTGTCCAGTC 2681
Qy 2161 GGAGGAGGAGCAGGAGGAGGCGCAGGGCTGAGTCCCAGTCCGAGGAGCAGCAGGAGGC 2220
Db 2682 AGAGGAGGAGCAGGAGGAGGCGCAGGGCTGAGTCCCAGTCCGAGGAGCAGCAGGAGGC 2741
Qy 2221 CAGGGCTCAGAGCCCACTGCCCCAGGTCAGTGCAGGCGCTGTGCTGAGGTCCGCGAGGCG 2280
Db 2742 CAGGGCTCAGAGCCCACTGCCCCAGGTCAGTGCAGGCGCTGTGCTGAGGTCCGCGAGGCG 2801
Qy 2281 TCCACACGAGGCTCTCCAGAGCCCACTGCCCCAGGTCAGTGCAGGCGCTGTGCTGAGGTCCGCGAGGCG 2861
Db 2802 TCCACACGAGGCTCTCCAGAGCCCACTGCCCCAGGTCAGTGCAGGCGCTGTGCTGAGGTCCGCGAGGCG 2861
Qy 2341 GCAGATCCGGGACCTGTGAGTGTGCGGAGGCGCGGACACATATCCCTGGACATTTTC 2400
Db 2862 GCAGATCCGGGACCTGTGAGTGTGCGGAGGCGCGGACACATATCCCTGGACATTTTC 2921
Qy 2401 CGAGGTGACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCATTT 2460
Db 2922 CGAGGTGACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCATTT 2981
Qy 2461 CGAGTTTATGATCTTACGAAAGTCCCAAGTCCGCTCAGCAGGCGCCCTCCCCCAT 2520
Db 2982 CGAGTTTATGATCTTACGAAAGTCCCAAGTCCGCTCAGCAGGCGCCCTCCCCCAT 3041
Qy 2521 GGCTGAGGAGAGCTGCGCGAGTTCCCGAGGCGCACGTCGCGCTGGCCAGGTGAATGGG 2580
Db 3042 GGCTGAGGAGAGCTGCGCGAGTTCCCGAGGCGCACGTCGCGCTGGCCAGGTGAATGGG 3101
Qy 2581 CCCCCACGAGGCTTGAGATCAGAGGAGTCAAGAGATGTGACGCGCTGCTGGGAGA 2640
Db 3102 CCCCCACGAGGCTTGAGATCAGAGGAGTCAAGAGATGTGACGCGCTGCTGGGAGA 3161
Qy 2641 GGCTGCGGTGGGAGGAGCGCAAGTGTCTCGCCGCTCAGCAGGCTCTTCCACTTCCC 2700
Db 3162 GGCTGCGGTGGGAGGAGCGCAAGTGTCTCGCCGCTCAGCAGGCTCTTCCACTTCCC 3221
Qy 2701 TGGAGGACCTGCGGCTGAGTACCTGACAGCTGGGCTGGCTGAGAGAGTGAAGGC 2760
Db 3222 TGGAGGACCTGCGGCTGAGTACCTGACAGCTGGGCTGGCTGAGAGAGTGAAGGC 3281
Qy 2761 CTCCGTGAGGACATCTCCCGATCTTGAAGGCGAGGCGGAGGCTGTGGAGAGGAGGG 2820
Db 3282 CTCCGTGAGGACATCTCCCGATCTTGAAGGCGAGGCGGAGGCTGTGGAGAGGAGGG 3341
Qy 2821 GCCCCCAGGAGGAGGAGCGGCTTGTCTTCCCTGCTCAGGTCTGAAGAGCTGGGA 2880
Db 3342 GCCCCCAGGAGGAGGAGCGGCTTGTCTTCCCTGCTCAGGTCTGAAGAGCTGGGA 3401
Qy 2881 CCGAGGCGGACATCTTAAAGGAGCTCTCAGATGAGACTGTGTCTGGGCCAGTCAGT 2940
Db 3402 CCGAGGCGGACATCTTAAAGGAGCTCTCAGATGAGACTGTGTCTGGGCCAGTCAGT 3461
Qy 2941 GACACTGCGCTGCGAGGTGTGAGGCGGAGCTGCGGAGGCGGAGGAGGAGGAGG 3000
Db 3462 GACACTGCGCTGCGAGGTGTGAGGCGGAGCTGCGGAGGCGGAGGAGGAGGAGG 3521

Qy 3001 A 3001
Db 3522 A 3522

RESULT 5

AAC62287
ID AAC62287 standard; cDNA; 7928 BP.

XX AAC62287;

XX AC

XX 19-MAR-2001 (first entry)

XX cDNA encoding a splice variant of a signal transduction polypeptide.

DE Signal transduction; H19G5; Kinase; cardiac disease; angina pectoris;
XX congestive heart failure; dilated congestive cardiomyopathy;
XX hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW mitral valve disease; aortic valve disease; tricuspid valve disease;
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW atherosclerosis; cardiac tumour; microbial infection; splice variant; as.

XX Homo sapiens.

XX Location/Qualifiers

XX Key 60..7850

XX CDS /*tag= a

XX FT /product= "signal transduction polypeptide H19G5 splice
XX variant"

XX WO200063381-A1.

XX 26-OCT-2000.

XX 11-APR-2000; 2000WO-US009488.

XX 16-APR-1999; 99US-0129553P.

XX (SCIO-) SCIOS INC.

XX Zeng W, Stanton L, Kong H;

XX WPI; 2001-007013/01.

XX P-PSDB; AAB30569.

XX Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction.

XX Claim 4; Page 65-68; 81pp; English.

PS The present sequence encodes a splice variant of human in signal
XX transduction polypeptide. The polypeptide is designated H19G5. The
XX protein is capable of regulating signal transduction and exhibits kinase
CC activity. The H19G5 transcript is expressed in the heart. H19G5
CC polypeptides and polynucleotides are useful for preventing or treating a
CC cardiac disease, such as congestive heart failure, dilated congestive
CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC mitral valve disease, aortic valve disease or tricuspid valve disease,
CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC arterial or rheovascular hypertension, arteriosclerosis, atherosclerosis
CC and cardiac tumours in humans. The polypeptide is also useful for
CC detecting the expression of a protein capable of regulating signal
CC transduction or the expression of a protein capable of acting as a donor
CC or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC be used as probes for detecting discrete antigens expressed by tissue or
CC cell samples, and therefore used in humans for localization and
XX monitoring of microbial infection

SQ Sequence 7928 BP; 1535 A; 2644 C; 2488 G; 1261 T; 0 U; 0 Other;

Query Match

99.9%; Score 2999.4; DB 5; Length 7928;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 3000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTACGGAGCAGAACTCGGGCCCGCAGGCATACAGGGAGCGAGACATCCCTGGCCGCGCTGAG 60
Db
3443 CCTACGGAGCAGAACTCGGGCCCGCAGGCATACAGGGAGCGAGACATCCCTGGCCGCGCTGAG 3502
QY 61 CCACCCGCTGGTCA CGGGCTCTGACCAAGCTTTGAGACCCCGCAAGACCTCATCTCAT 120
Db 3503 CCACCCGCTGGTCA CGGGCTCTGACCAAGCTTTGAGACCCCGCAAGACCTCATCTCAT 3562
QY 121 CCTGAGCTGTGCTCATCCGAGAGCTGTGACCGCTGTACAGACCGCTGTACAGAAAGGGGCTGGTAC 180
Db 3563 CCTGAGCTGTGCTCATCCGAGAGCTGTGACCGCTGTACAGAAAGGGGCTGGTAC 3622
QY 181 GGAGGCCGAGGTCAAGGCTTACATCCAGCAGCTGTGAGGGGCTGCACTACCTTGCACAG 240
Db 3623 GGAGGCCGAGGTCAAGGCTTACATCCAGCAGCTGTGAGGGGCTGCACTACCTTGCACAG 3682
QY 241 CCATGGGCTTTCACACCTGGACATAAAGCCCTTAACATCTCTGATGGTGCATCTCTGCCG 300
Db 3683 CCATGGGCTTTCACACCTGGACATAAAGCCCTTAACATCTCTGATGGTGCATCTCTGCCG 3742
QY 301 GGAAGACATTAATACTGGACTTTGGCTTTGGCTTTGCCAGAAATCAACCCAGCAGAGCTGCA 360
Db 3743 GGAAGACATTAATACTGGACTTTGGCTTTGGCTTTGCCAGAAATCAACCCAGCAGAGCTGCA 3802
QY 361 GTTCAGCCAGTACGGCTCCCTCGATTCGTCTGATTCCTTAACATCTCTGATGGTGCATCTCTGCCG 420
Db 3803 GTTCAGCCAGTACGGCTCCCTCGATTCGTCTGATTCCTTAACATCTCTGATGGTGCATCTCTGCCG 3862
QY 421 GAGGAAAGCTCCGACATTTGGGCAATGGGTGTGATCTCTTACCTCAGCTGACCTGCTC 480
Db 3863 GAGGAAAGCTCCGACATTTGGGCAATGGGTGTGATCTCTTACCTCAGCTGACCTGCTC 3922
QY 481 ATCCCCATTTCCCGCAGAGTGACCGTGCCACCTCTCGAACCTCTCGAACCTCTCGAACCGGCGGT 540
Db 3923 ATCCCCATTTCCCGCAGAGTGACCGTGCCACCTCTCGAACCTCTCGAACCTCTCGAACCGGCGGT 3982
QY 541 GTCATGAGCAGCCCCATGGTGTGCCACCTCAGCGAAGAGCGCCAAAGACTTTCATCAAGGC 600
Db 3983 GTCATGAGCAGCCCCATGGTGTGCCACCTCAGCGAAGAGCGCCAAAGACTTTCATCAAGGC 4042
QY 601 TACGCTCAGAGAGCCCTCAGGCCCGGCTAGTGGCGGCCAGTGCCTCTCCACCCCTG 660
Db 4043 TACGCTCAGAGAGCCCTCAGGCCCGGCTAGTGGCGGCCAGTGCCTCTCCACCCCTG 4102
QY 661 GTTCTGAAATCCATGCTCTGCGAGGAGGCCACTTCATCAACACCAAGCAGCTCAAGTT 720
Db 4103 GTTCTGAAATCCATGCTCTGCGAGGAGGCCACTTCATCAACACCAAGCAGCTCAAGTT 4162
QY 721 CTTCTGGCCGAAAGTCTGCTGGCAGCTTCCCTGATGAGCTACAAGTCCAATCTCTGGTAT 780
Db 4163 CTTCTGGCCGAAAGTCTGCTGGCAGCTTCCCTGATGAGCTACAAGTCCAATCTCTGGTAT 4222
QY 781 GCGCTCATCCCTGAGCTGTGGGGCCCAACCGCAGAGCCCTCCCTGGCGGTAGCCG 840
Db 4223 GCGCTCATCCCTGAGCTGTGGGGCCCAACCGCAGAGCCCTCCCTGGCGGTAGCCG 4282
QY 841 GCACCTCTGAGGAGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTCTGACAAACGAGCT 900
Db 4283 GCACCTCTGAGGAGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTCTGACAAACGAGCT 4342
QY 901 CGGCCCATTTGCCGGGCTTAAGTCACTGCCACCTCCCGGTGACACACTCACTCACTGCT 960
Db 4343 CGGCCCATTTGCCGGGCTTAAGTCACTGCCACCTCCCGGTGACACACTCACTCACTGCT 4402
QY 961 GCACCCCGGGGCTTCTCTGGGCCCTCGCCAGCTGCTGAGGAGCGAGGCGAGTGA 1020
Db 4403 GCACCCCGGGGCTTCTCTGGGCCCTCGCCAGCTGCTGAGGAGCGAGGCGAGTGA 4462
QY 1021 GCGCTCAACGAGGCCCAAGCTCCGCTGCACTCTCCGAGGGGTGCGGGCCACCGGCGC 1080

Db 4463 GCGCTCAACGAGGCCCAAGCTCCGCTGCACTCTCCGAGGGGTGCGGGCCACCGGCGC 4522
QY 1081 CCAGGGCTGCGTGTCCCGGCAACAGGCTCATCCGAGGCTGTTCTACCAACAGGCGGGTGA 1140
Db 4523 CCAGGGCTGCGTGTCCCGGCAACAGGCTCATCCGAGGCTGTTCTACCAACAGGCGGGTGA 4582
QY 1141 GAGCCCTTAGACACAGGGGCTCTGGCCCGGGGAGCAGGGGCAACCGGCGCGGCGGCA 1200
Db 4583 GAGCCCTTAGACACAGGGGCTCTGGCCCGGGGAGCAGGGGCAACCGGCGCGGCGGCA 4642
QY 1201 CTTGCTGAAGGGGGTACATTTGCGGGGGCTGCTGCCAGGCTTGCGCAGACACTGATGGA 1260
Db 4643 CTTGCTGAAGGGGGTACATTTGCGGGGGCTGCTGCCAGGCTTGCGCAGACACTGATGGA 4702
QY 1261 GCACCGGCTGCTGGAGAGAGGCGCGCAGGAGAGGAGCAGGCAACCTCTCTGGCCAAAGC 1320
Db 4703 GCACCGGCTGCTGGAGAGAGGCGCGCAGGAGAGGAGCAGGCAACCTCTCTGGCCAAAGC 4762
QY 1321 CCCCTCATTTGAGACTGCGCTCTCGGCTGCTCTGGCACCACCTTTGGCCCTTGSCCA 1380
Db 4763 CCCCTCATTTGAGACTGCGCTCTCGGCTGCTCTGGCACCACCTTTGGCCCTTGSCCA 4822
QY 1381 CAGCACTCCCTGAAACATGACTCTCCGAGCACCCCGCCCTCTCTCGAGGCTGCGG 1440
Db 4823 CAGCACTCCCTGAAACATGACTCTCCGAGCACCCCGCCCTCTCTCGAGGCTGCGG 4882
QY 1441 TGAGGCACAGCACTGCTTCAGCCCTCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1500
Db 4883 TGAGGCACAGCACTGCTTCAGCCCTCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 4942
QY 1501 CCCTCAGGGTCCAAAGCAGCTTCCATCTCAGTGGGCGCACCCAGGCACTGCTCAGCCGA 1560
Db 4943 CCCTCAGGGTCCAAAGCAGCTTCCATCTCAGTGGGCGCACCCAGGCACTGCTCAGCCGA 5002
QY 1561 GAGGCCATCCCGGACAGCCCTTGGGGGAGCAGCCCTTTCTGCGCAACCCCAAGCAGG 1620
Db 5003 GAGGCCATCCCGGACAGCCCTTGGGGGAGCAGCCCTTTCTGCGCAACCCCAAGCAGG 5062
QY 1621 TTCTGCCCCCAGGAGGGCTGACGCCCCCAGCCAGCAGTTGCCCATGCCCTCTGGCTC 1680
Db 5063 TTCTGCCCCCAGGAGGGCTGACGCCCCCAGCCAGCAGTTGCCCATGCCCTCTGGCTC 5122
QY 1681 CTTTCTCAGGATCTTGCAAGAGGCCCTTTAGTACCTCAAGCCCTTTCTTGGGACA 1740
Db 5123 CTTTCTCAGGATCTTGCAAGAGGCCCTTTAGTACCTCAAGCCCTTTCTTGGGACA 5182
QY 1741 GCGCCAGGACCCCTGCGCCCTGCCAAGCAAGCCCCCATTGGACTCTAAGATGGGGC 1800
Db 5183 GCGCCAGGACCCCTGCGCCCTGCCAAGCAAGCCCCCATTGGACTCTAAGATGGGGC 5242
QY 1801 TGGAGACATCTCTCTCTGGGAGGCCAAACCCCGGCCCTCAGTTCCCGAGGGTCAAG 1860
Db 5243 TGGAGACATCTCTCTCTGGGAGGCCAAACCCCGGCCCTCAGTTCCCGAGGGTCAAG 5302
QY 1861 CTTCCAGGCGAGCTTTCCCAAGTGAAGTCTCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 1920
Db 5303 CTTCCAGGCGAGCTTTCCCAAGTGAAGTCTCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 5362
QY 1921 AGAGCTGGCCCTCTCTGGATCGGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCA 1980
Db 5363 AGAGCTGGCCCTCTCTGGATCGGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCA 5422
QY 1981 CTTCCACACCCACTTTGACGCGGCTCAGAAACAGGCGACCATGCCAAAGTTCTCCCTGG 2040
Db 5423 CTTCCACACCCACTTTGACGCGGCTCAGAAACAGGCGACCATGCCAAAGTTCTCCCTGG 5482
QY 2041 TGGTCCGGGGGCTTACGACGCGTGGCTGTGCTATGGCACTTTGCTTTTGGTGGAGATGC 2100
Db 5483 TGGTCCGGGGGCTTACGACGCGTGGCTGTGCTATGGCACTTTGCTTTTGGTGGAGATGC 5542
QY 2101 AGGGGCGATGCTGGGGCAGGGGCCCATGTGGGCCAGGATAGCCTGGGCTGTGCCAGTC 2160
Db 5543 AGGGGCGATGCTGGGGCAGGGGCCCATGTGGGCCAGGATAGCCTGGGCTGTGCCAGTC 5602

Qy	2161	GGAGGAGGAGCAGCAGGAGGAGCCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC	2220
Db	5603	AGAGGAGGAGCAGCAGGAGGAGGCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC	5662
Qy	2221	CAGGGCTGAGAGCCACTGCCCCAGGTCAGTGCAAGGCCTGTGCCTCAGGTCGCGCAGGGC	2280
Db	5663	CAGGGCTGAGAGCCCACTGCCCCAGGTCAGTGCAAGGCCTGTGTCTCAGGTCGCGCAGGGC	5722
Qy	2281	TCCACACGAGGAGCTCTCCAGAGCCACCCTCATGGGAGGACATCGGGCAGGTCCTCCCTGGT	2340
Db	5723	TCCACACGAGGAGCTCTCCAGAGCCACCCTCATGGGAGGACATCGGGCAGGTCCTCCCTGGT	5782
Qy	2341	GCAGATCGGGACCTGTTCAGGTGATGCGGAGCGGCGGACACAAATATCCCTTGACATTTTC	2400
Db	5783	GCAGATCGGGACCTGTTCAGGTGATGCGGAGCGGCGGACACAAATATCCCTTGACATTTTC	5842
Qy	2401	CGAGGTGGACCCCGCCCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTTCCCAT	2460
Db	5843	CGAGGTGGACCCCGCCCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTTCCCAT	5902
Qy	2461	CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCCAT	2520
Db	5903	CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCCAT	5962
Qy	2521	GGCTGAGGAGGAGCTGGCCGAGTTCCCGGAGCCCACTGGCCCTGGGCCAGGTGACTGGG	2580
Db	5963	GGCTGAGGAGGAGCTGGCCGAGTTCCCGGAGCCCACTGGCCCTGGGCCAGGTGACTGGG	6022
Qy	2581	CCCCACGCGAGGCTTGAGATTCACAGAGGAGTCAAGAGGATGTGGACGCGTGTCTGGCAGA	2640
Db	6023	CCCCACGCGAGGCTTGAGATTCACAGAGGAGTCAAGAGGATGTGGACGCGTGTCTGGCAGA	6082
Qy	2641	GGCTGCGTGGCAGGAGCGCAAGTGTCTTCGCGCTCAGCAGCGCTTCCACTTTCCC	2700
Db	6083	GGCTGCGTGGCAGGAGCGCAAGTGTCTTCGCGCTCAGCAGCGCTTCCACTTTCCC	6142
Qy	2701	TGGGAGGACCTGCGCTGGATGAGCTGCAGAGCTGGGGCTGGTGAGAGTGAAGGC	2760
Db	6143	TGGGAGGACCTGCGCTGGATGAGCTGCAGAGCTGGGGCTGGTGAGAGTGAAGGC	6202
Qy	2761	CTCCGTGGAGCACATCTCCCGATCTCTGAAGGGCAGGCCGGAAGTCTGGAGAGGAGGG	2820
Db	6203	CTCCGTGGAGCACATCTCCCGATCTCTGAAGGGCAGGCCGGAAGTCTGGAGAGGAGGG	6262
Qy	2821	GCCCCCAGGAAGAGCAGGCTTGTCTTCCGGCTCTCAGGTCGGAAGAGCTGGGA	2880
Db	6263	GCCCCCAGGAAGAGCAGGCTTGTCTTCCGGCTCTCAGGTCGGAAGAGCTGGGA	6322
Qy	2881	CCGAGCCCGACATTTCTTAAAGGAGCTCTCAGATGAGACTGTGGTCTGGGCCAGTCACT	2940
Db	6323	CCGAGCCCGACATTTCTTAAAGGAGCTCTCAGATGAGACTGTGGTCTGGGCCAGTCACT	6382
Qy	2941	GACACTGGCCTGCCAGGTGTTCAGGCCAGCAGCTGCCAGGCCACTCTGGAGCAAGACGG	3000
Db	6383	GACACTGGCCTGCCAGGTGTTCAGGCCAGCAGCTGCCAGGCCACTCTGGAGCAAGACGG	6442
Qy	3001	A 3001	
Db	6443	A 6443	

RESULT 6
AAI-43909

AAL43909
 ID AAL43909 standard; DNA: 5207 BP.

AC AAL43909;

DT 19-SEP-2002 (first entry)

XX Human kinase protein coding sequence 2.

Human; gene; ds; gene therapy; chromosome 1; kinase protein;

myosin light chain kinase subfamily; kinase protein-mediated disease; transgenic animal.

Homo sapiens.

Key	Location/Qualifiers
CDS	109. .5106

```

variation
/product= "Human kinase protein"
replace(311, C)
/ tag= a

```

```
variation
/standard name= "Single nucleotide polymorphism"
/label= C
replace(311, T)
```

```

variation
/ - tag= D
/standard_name= "Single nucleotide polymorphism"
replace(1741, T)

```

```
variation
/standard_name= "Single nucleotide polymorphism"
replace(2714, C)
```

```
variation
/*tag= e
/standard_name= "Single nucleotide polymorphism".
replace(2745, T)
```

```
variation
/ntag= 1
/standard name= "Single nucleotide polymorphism"
replace(2859, A)
```

```
variation
/*tag= g
/standard name= "Single nucleotide polymorphism"
replace(3420, C)
```

```

/ *tag= n
/standard_name= "Single nucleotide polymorphism"

```

WO200240683-A2.

23-MAY-2002.

22-OCT-2001: 2001WO-US032616.

14-NOV-2000: 2000US-00711134.

17-MAY-2001; 2001US-00858664.

(PEKE) PE CORP NY.

Wei M, Ketchum K, Di Francesco V, Beasley EM;

WPI: 2002-500223/53.

P-PSDB; AAO15372.

New kinase proteins related to myosin light chain kinase subfamily and encoding polynucleotide, useful for diagnosing, treating disease or condition mediated by the kinase protein and for identifying modulators.

Claim 23: Fig 3: 96pp: English.

The invention comprises the amino acid and coding sequences (located on chromosome 1) of a human kinase protein that is related to the myosin light chain kinase subfamily. The human kinase DNA and protein sequences of the invention are useful for identifying agents that modulate the activity of the human kinase protein. Kinase-modulating agents are useful for treating a disease or condition mediated by a human kinase protein. The human kinase DNA sequences can be used to produce transgenic animals, which are useful for studying the function of kinase proteins and identifying/evaluating modulators of kinase protein activity. The present DNA sequence encodes the human kinase protein of the invention.

Sequence 5207 BP: 988 A: 1758 C: 1612 G: 849 T: 0 U: 0 Other:

Query Match 99.9%; Score 2997.8; DB 6; Length 5207;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

10y

Db 699

699 CCTACGGAGCAGAACTCGGGCCAGGCATACAGGGAGCGAGACATCCTGGCCGCGTGAG 758

QY 61 CCACCCGCTGGTCA CCGGGCTCTCTGGACCAAGTTTGAGACCCCGAAGACCCCTCATCTCAT 120
DB |||||
759 CCACCCGCTGGTCA CCGGGCTCTCTGGACCAAGTTTGAGACCCCGAAGACCCCTCATCTCAT 818
QY 121 CCTGGAGCTGTCTCAT CCGAGAGCTGTGTGACCCCTGTATACAGGAAGGGCGTGGTGAC 180
DB |||||
819 CTTGGAGCTGTCTCAT CCGAGAGCTGTGTGACCCCTGTATACAGGAAGGGCGTGGTGAC 878
QY 181 GGAGGCCGAGGTCAAGGTCTACATCCAGCAGCTGTGTGAGGGGTGTCACTACCTGACAG 240
DB |||||
879 GGAGGCCGAGGTCAAGGTCTACATCCAGCAGCTGTGTGAGGGGTGTCACTACCTGACAG 938
QY 241 CCATGGCGTTCTTCCA CTTGGACATTAAGCCCTCTTAACATCTGTATGGTGTGCTGCCCCG 300
DB |||||
939 CCATGGCGTTCTTCCA CTTGGACATTAAGCCCTCTTAACATCTGTATGGTGTGCTGCCCCG 998
QY 301 GGAAAGACATTAATCTCGACTTTGGCTTTGGCCAGAAACATCAACCCAGCAGAGCTGCA 360
DB |||||
999 GGAAGACATTAATCTCGACTTTGGCTTTGGCCAGAAACATCAACCCAGCAGAGCTGCA 1058
QY 361 GTTCAGCCAGTACCGCTCCCTGAGTTTGGTCTCTCCCGAGATCATCCAGCAGAAACCCCTGT 420
DB |||||
1059 GTTCAGCCAGTACCGCTCCCTGAGTTTGGTCTCTCCCGAGATCATCCAGCAGAAACCCCTGT 1118
QY 421 GAGCGAAGCCTCCGACATTTGGGCGATGGGTGTATCTCTACCTCAGCTGACCTGCTC 480
DB |||||
1119 GAGCGAAGCCTCCGACATTTGGGCGATGGGTGTATCTCTCTCTACCTCAGCTGACCTGCTC 1178
QY 481 ATCCCCATTTGCCCGGAGAGTGACCGGTGCCACCTCTCTGAACCTCTCTGAGGGGGCGGT 540
DB |||||
1179 ATCCCCATTTGCCCGGAGAGTGACCGGTGCCACCTCTCTGAACCTCTCTGAGGGGGCGGT 1238
QY 541 GTCATGAGAGAGCCCCATTTGGTGTCCCACTCAGCGAAGACGCCAAAGACTTCATCAAGGC 600
DB |||||
1239 GTCATGAGAGAGCCCCATTTGGTGTCCCACTCAGCGAAGACGCCAAAGACTTCATCAAGGC 1298
QY 601 TACGCTCAGAGAGCCCTCAGCGCGGCTTAGTGTGGCGCCGAGTGCCTCTCCACCCCTG 660
DB |||||
1299 TACGCTCAGAGAGCCCTCAGCGCGGCTTAGTGTGGCGCCGAGTGCCTCTCCACCCCTG 1358
QY 661 GTTCCTGAAATCCATGCTGCGGAGAGGGCCACTTCATCAACCAACCAAGCAGCTCAAGTT 720
DB |||||
1359 GTTCCTGAAATCCATGCTGCGGAGAGGGCCACTTCATCAACCAACCAAGCAGCTCAAGTT 1418
QY 721 CCTCCTGGCGGAAGTCTGCGAGCGTTCCCTGTATGAGCTACAAGTCCATCTGTGTGAT 780
DB |||||
1419 CCTCCTGGCGGAAGTCTGCGAGCGTTCCCTGTATGAGCTACAAGTCCATCTGTGTGAT 1478
QY 781 GCGCTCCATCCCTGAGCTGTGCGGGGCCACCGCAGAGCCCTCCCTCGGCGTAGCCG 840
DB |||||
1479 GCGCTCCATCCCTGAGCTGTGCGGGGCCACCGCAGAGCCCTCCCTCGGCGTAGCCG 1538
QY 841 GCACCTCTGAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTCTCTCTGACCAACGAGCT 900
DB |||||
1539 GCACCTCTGAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTCTCTCTGACCAACGAGCT 1598
QY 901 CGCCCCATTTGCCGGGCTAAGTCACTGCCACCTCCCGGTTGACACACTACCACTGTCT 960
DB |||||
1599 CGCCCCATTTGCCGGGCTAAGTCACTGCCACCTCCCGGTTGACACACTACCACTGTCT 1658
QY 961 GCACCCCGGGGCTTCTGCGGGCTCTCGGCGCCCTGCGCTGAGGAAGCCGAGGCCAGTGCA 1020
DB |||||
1659 GCACCCCGGGGCTTCTGCGGGCTCTCGGCGCCCTGCGCTGAGGAAGCCGAGGCCAGTGCA 1718
QY 1021 GCGCTTCAACGAGGCCCTGAGCTCCGCTGTATCTCCGAGGGTGTCCGGGCCACCGGCCGC 1080
DB |||||
1719 GCGCTTCAACGAGGCCCTGAGCTCCGCTGTATCTCCCGAGGGTGTCCGGGCCACCGGCCGC 1778
QY 1081 CCAGGGCTGTGCTCCCGGCACAGCGTTCATCCGACGCTGTCTACCAACGAGCGGGTCA 1140
DB |||||
1779 CCAGGGCTGTGCTCCCGGCACAGCGTTCATCCGACGCTGTCTACCAACGAGCGGGTCA 1838

QY 1141 GAGCCCTGAGCAGCGGGCCCTTGGCCCCCGGGAGCAGCGCGCACCCGCGCGCGCGCGCA 1200
DB |||||
1839 GAGCCCTGAGCAGCGGGCCCTTGGCCCCCGGGAGCAGCGCGCACCCGCGCGCGCGCGCA 1898
QY 1201 CTTGCTAAAGGGCGGTACATTTGGGGGGCGCTGCCAGGCTGTGCGAGCCACTGTATGGA 1260
DB |||||
1899 CTTGCTAAAGGGCGGTACATTTGGGGGGCGCTGCCAGGCTGTGCGAGCCACTGTATGGA 1958
QY 1261 GCACCGGCTCTGGAGGAGGAGGGCGCCAGGGAGGAGCAGGCCACCTCTCTGGCCAAAGC 1320
DB |||||
1959 GCACCGGCTCTGGAGGAGGAGGGCGCCAGGGAGGAGCAGGCCACCTCTCTGGCCAAAGC 2018
QY 1321 CCCCCTATTGAGACTGCCCTCCGGCTGTCTGTGACCAACCTGTGGCCCTCTGCGCA 1380
DB |||||
2019 CCCCCTATTGAGACTGCCCTCCGGCTGTCTGTGACCAACCTGTGGCCCTCTGCGCA 2078
QY 1381 CAGGCACTCCTTGAAATGATCTCTGAGCAACCCCGCCCTCTCTGAGGCTGTGCG 1440
DB |||||
2079 CAGGCACTCCTTGAAATGATCTCTGAGCAACCCCGCCCTCTCTGAGGCTGTGCG 2138
QY 1441 TGAGGCAACAGCACTGCTTTCAGCCCTCCGGGGGGGCCCTTATCAGGACATGGGGCA 1500
DB |||||
2139 TGAGGCAACAGCACTGCTTTCAGCCCTCCGGGGGGGCCCTTATCAGGACATGGGGCA 2198
QY 1501 CCTCAGGGCTCCAGCAGCTTCCATCCATCTGTGTGGCCACCCAGGCACTGCTCAGCCAGA 1560
DB |||||
2199 CCTCAGGGCTCCAGCAGCTTCCATCCATCTGTGTGGCCACCCAGGCACTGCTCAGCCAGA 2258
QY 1561 GAGGCCATCCCGGACAGCCCTTGGGGGGCAGCCAGCCCTTCTGTGCAACCCCAAGCAGGG 1620
DB |||||
2259 GAGGCCATCCCGGACAGCCCTTGGGGGGCAGCCAGCCCTTCTGTGCAACCCCAAGCAGGG 2318
QY 1621 TTCTGCCCCCAGGAGGGCTGAGCCCCCAGCAGTTGCCCCATGCCCCCTCTCTGCTC 1680
DB |||||
2319 TTCTGCCCCCAGGAGGGCTGAGCCCCCAGCAGTTGCCCCATGCCCCCTCTCTGCTC 2378
QY 1681 CTTTCTCTCAGGATCTTTGCAAGAGCCCTTGTATACCTCAAGCCCTTCTTGGGACA 1740
DB |||||
2379 CTTTCTCTCAGGATCTTTGCAAGAGAGCCCTTGTATACCTCAAGCCCTTCTTGGGACA 2438
QY 1741 GCCCAGGACACCCCTGCCCCCTGCCAAGCAAGCCCCCATTTGGAATTAAGTGGGGC 1800
DB |||||
2439 GCCCAGGACACCCCTGCCCCCTGCCAAGCAAGCCCCCATTTGGAATTAAGTGGGGC 2498
QY 1801 TGGAGACATCTCTTCTTCTGGAGGCGCAAAACCCCGGCCCTCGAGTTCCCCAGGGTCAAG 1860
DB |||||
2499 TGGAGACATCTCTTCTTCTGGAGGCGCAAAACCCCGGCCCTCGAGTTCCCCAGGGTCAAG 2558
QY 1861 CTTCCAGGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 1920
DB |||||
2559 CTTCCAGGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 2618
QY 1921 AGAGCTGCGCCCTCCCTGGATGCGGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 1980
DB |||||
2619 AGAGCTGCGCCCTCCCTGGATGCGGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 2678
QY 1981 CTTCCAACCCACTTTGACGCGGCTCAGGAAACAGCGCACCATGCGCAAGTTCTCCCTGGG 2040
DB |||||
2679 CTTCCAACCCACTTTGACGCGGCTCAGGAAACAGGTGACCATGCGCAAGTTCTCCCTGGG 2738
QY 2041 TGGTCGCGGGGGCTACGAGGGGTGGCTGTATGGCACTTTGCTTTGTTGGAGATGC 2100
DB |||||
2739 TGGTCGCGGGGGCTACGAGGGGTGGCTGTATGGCACTTTGCTTTTGGTGGAGATGC 2798
QY 2101 AGGGGCACTGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGTCCAGTTC 2160
DB |||||
2799 AGGGGCACTGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGTCCAGTTC 2858
QY 2161 GGAGGAGGAGGACGAGGAGGAGGCGCGGCTGATGTTCCAGTCTGGAGGAGCAGCAGAGGC 2220
DB |||||
2859 GGAGGAGGAGGACGAGGAGGAGGCGCGGCTGATGTTCCAGTCTGGAGGAGCAGCAGAGGC 2918
QY 2221 CAGGGCTGAGAGCCCACTGCCCCAGGTGAGTCAAGGGCTGTGCTGAGGTCGGCAGGGC 2280

```
Db 2919 CAGGGCTGAGAGCCCACTGCCCGAGGTGAGTGAAGGCTGTGCTGAGGTGCGAGGGC 2978
Qy 2281 TCCACACGAGAGCTCTCCAGAGCCCAACCCATGAGGAGACATCGGGAGGTCTCCCTGGT 2340
Db 2979 TCCACACGAGAGCTCTCCAGAGCCCAACCCATGAGGAGACATCGGGAGGTCTCCCTGGT 3038
Qy 2341 GCAGATCCGGAGCCTGTCAAGTGTGCGGAGCGCCGACACAATATCCCTGGACATTC 2400
Db 3039 GCAGATCCGGAGCCTGTCAAGTGTGCGGAGCGCCGACACAATATCCCTGGACATTC 3098
Qy 2401 CGAGGTGAGACCCCGCTACCTCAACCTCTCAGAGCTGTACGATATCAAGTACCTCCCAT 2460
Db 3099 CGAGGTGAGACCCCGCTACCTCAACCTCTCAGAGCTGTACGATATCAAGTACCTCCCAT 3158
Qy 2461 CGAGTTTATGATCTTCAGGAAGTTCCTGAGAGTCCGCTCAGCAGAGCGCCCTCCCCAT 2520
Db 3159 CGAGTTTATGATCTTCAGGAAGTTCCTGAGAGTCCGCTCAGCAGAGCGCCCTCCCCAT 3218
Qy 2521 GGCTGAGGAGAGCTGCCGAGTTCCCGGAGCCCACTGCGCTGCGCCCTGCGAGTGAAGTGG 2580
Db 3219 GGCTGAGGAGAGCTGCCGAGTTCCCGGAGCCCACTGCGCTGCGCCCTGCGAGTGAAGTGG 3278
Qy 2581 CCCCACGCGAGCCTGGAGATCACAGAGAGTCAAGAGTGTGACGCGCTGTGGGAGA 2640
Db 3279 CCCCACGCGAGCCTGGAGATCACAGAGAGTCAAGAGTGTGACGCGCTGTGGGAGA 3338
Qy 2641 GGCTGCGCTGGCAGGAGCGCAAGTGTCTCCGCGTCAAGAGCTCTTCCACTTCCC 2700
Db 3339 GGCTGCGCTGGCAGGAGCGCAAGTGTCTCCGCGTCAAGAGCTCTTCCACTTCCC 3398
Qy 2701 TGGAGGACCTCGCTGATGAGCTGACAGAGTGGGCTGCTGAGAGTGAAGGC 2760
Db 3399 TGGAGGACCTCGCTGATGAGCTGACAGAGTGGGCTGCTGAGAGTGAAGGC 3458
Qy 2761 CTCCGTGAGACATCTCCCGATCTCAGAGGAGCGGAGGCTCTGAGAGAGGAGG 2820
Db 3459 CTCCGTGAGACATCTCAGAGGAGCGGAGGCTCTGAGAGAGGAGG 3518
Qy 2821 GCGCCCGCAGGAGAGCGAGCTTGTCTTCCGCTCTCAGGTCTGAAGAGCTGGGA 2880
Db 3519 GCGCCCGCAGGAGAGCGAGCTTGTCTTCCGCTCTCAGGTCTGAAGAGCTGGGA 3578
Qy 2881 CCGAGCCCGACATCTCAAGGAGCTCTCAGATGAGACTGTGTCTGCGCCAGTCAGT 2940
Db 3579 CCGAGCCCGACATCTCAAGGAGCTCTCAGATGAGACTGTGTCTGCGCCAGTCAGT 3638
Qy 2941 GACACTGCGCTGCGAGGTGTGAGCCAGCAGCTGCCAGCCACCTGGAGCAAGAGCG 3000
Db 3639 GACACTGCGCTGCGAGGTGTGAGCCAGCAGCTGCCAGCCACCTGGAGCAAGAGCG 3698
```

RESULT 7

AAL43908

ID AAL43908 standard; DNA; 5207 BP.

XX AAL43908;

AC AAL43908;

XX AAL43908;

DT 19-SEP-2002 (first entry)

XX Human kinase protein coding sequence 1.

XX Human; gene; ds; gene therapy; chromosome 1; kinase protein;

KW myosin light chain kinase subfamily; kinase protein-mediated disease;

XX transgenic animal.

XX Homo sapiens.

OS Homo sapiens.

XX Key

PH Location/Qualifiers

```
FT CDS 109..5106
FT /*tag=a
FT /product= "Human kinase protein"
XX WO200240683-A2.
XX 23-MAY-2002.
XX 22-OCT-2001; 2001WO-US032616.
XX 14-NOV-2000; 2000US-00711134.
XX 17-MAY-2001; 2001US-00858664.
XX (PEKE ) PE CORP NY.
XX Wei M, Ketchum K, Di Francesco V, Beasley EM;
XX WPI; 2002-500223/53.
XX P-PSDB; AAO15372.
XX New kinase proteins related to myosin light chain kinase subfamily and
XX encoding polynucleotide, useful for diagnosing, treating disease or
XX condition mediated by the kinase protein and for identifying modulators.
XX Claim 23; Fig 1; 96pp; English.
XX The invention comprises the amino acid and coding sequences (located on
XX chromosome 1) of a human kinase protein that is related to the myosin
XX light chain kinase subfamily. The human kinase DNA and protein sequences
XX of the invention are useful for identifying agents that modulate the
XX activity of the human kinase protein. Kinase-modulating agents are useful
XX for treating a disease or condition mediated by a human kinase protein.
XX The human kinase DNA sequences can be used to produce transgenic animals
XX which are useful for studying the function of kinase proteins and
XX identifying/evaluating modulators of kinase protein activity. The present
XX DNA sequence encodes the human kinase protein of the invention.
XX SQ Sequence 5207 BP; 988 A; 1758 C; 1612 G; 849 T; 0 U; 0 Other;
XX Query Match 99.9%; Score 2997.8; DB 6; Length 5207;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CCTACGAGCAGAGAACTCGGGCCAGGATACAGGAGGAGGAGACATCTGGCCGCGCTGAG 60
Db 699 CCTACGAGCAGAGAACTCGGGCCAGGATACAGGAGGAGGAGACATCTGGCCGCGCTGAG 758
Qy 61 CCACCCGCTGTACGCGGGCTGTGGACCAAGTTTGTAGACCGCAAGACCCCTCATCTCAT 120
Db 759 CCACCCGCTGTACGCGGGCTGTGGACCAAGTTTGTAGACCGCAAGACCCCTCATCTCAT 818
Qy 121 CCTGGAGCTGTGCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAGGCGGTGGTGAC 180
Db 819 CCTGGAGCTGTGCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAGGCGGTGGTGAC 878
Qy 181 GGAGCCGAGGTCAAGGTCTACATCCAGAGCTGTGGAGGGGCTGCACTACCTGCAAG 240
Db 879 GGAGCCGAGGTCAAGGTCTACATCCAGAGCTGTGGAGGGGCTGCACTACCTGCAAG 938
Qy 241 CCATGGGCTTCTCCACCTGGACATAAAGCCCTTAACATCTGTATGTGTGATCTCTGCCCG 300
Db 939 CCATGGGCTTCTCCACCTGGACATAAAGCCCTTAACATCTGTATGTGTGATCTCTGCCCG 998
Qy 301 GGAAGACATTAAATCTGCGACTTTGGCTTTGGCCAGAAACATCAACCCAGCAGAGTGA 360
Db 999 GGAAGACATTAAATCTGCGACTTTGGCTTTGGCCAGAAACATCAACCCAGCAGAGTGA 1058
Qy 361 GTTCAGCAGTACGGCTCCCTGTGATGCTGCTCCCGAGATCATCCAGCAGAGACCTGT 420
Db 1059 GTTCAGCAGTACGGCTCCCTGTGATGCTGCTCCCGAGATCATCCAGCAGAGACCTGT 1118
Qy 421 GAGCGAGACCTCCGACATTTGGGCGCATGGGTGTCTCTTACCTACGCTGAGCTGCTC 480
```

Db 1119 GAGCGAAGCCTCCGACATTTGGGCCAATGGGTGTCTCTTACCTCAGCTGACCTGTCTC 1178
QY 481 ATCCCAATTTGCGGCGAGGTGAACCGTGCACACCTCTCTGAACGTCCTTGGAGGGCGCGT 540
Db 1179 ATCCCAATTTGCGGCGAGGTGAACCGTGCACACCTCTCTGAACGTCCTTGGAGGGCGCGT 1238
QY 541 GTCATGGAGACGCCCATGGCTGGCCACCTCAGCGAAGACGCGCAAGACTTTCATCAAGGC 600
Db 1239 GTCATGGAGACGCCCATGGCTGGCCACCTCAGCGAAGACGCGCAAGACTTTCATCAAGGC 1298
QY 601 TACGCTGCAGAGAGCCCTCAGGCCCGGCGCTAGTGGGCGCCAGTGCCCTCTCCACCCCTG 660
Db 1299 TACGCTGCAGAGAGCCCTCAGGCCCGGCGCTAGTGGGCGCCAGTGCCCTCTCCACCCCTG 1358
QY 661 GTTCTGAAATCCATGCTGCGGAGGAGGCCACATTCATCAACACCAAGCAGCTCAAGTT 720
Db 1359 GTTCTGAAATCCATGCTGCGGAGGAGGCCACATTCATCAACACCAAGCAGCTCAAGTT 1418
QY 721 CCTCTGCGCCGAGTCTGCTGGCAGCGTTCCCTGANTGAGCTACAAAGTCCATCTCTGGTGA 780
Db 1419 CCTCTGCGCCGAGTCTGCTGGCAGCGTTCCCTGANTGAGCTACAAAGTCCATCTCTGGTGA 1478
QY 781 GCGCTCCATCCTTAGCTGCTGGGGCCCAACCGACAGCCCTCCCTCGGCGTAGCCCG 840
Db 1479 GCGCTCCATCCTTAGCTGCTGGGGCCCAACCGACAGCCCTCCCTCGGCGTAGCCCG 1538
QY 841 GCACTCTGCGAGGACACTGTGTGCTCCTCCAGTTCTCTCTCTCTCTGACAAACAGAGCT 900
Db 1539 GCACTCTGCGAGGACACTGTGTGCTCCTCCAGTTCTCTCTCTCTCTGACAAACAGAGCT 1598
QY 901 CGCCCCATTTGCGCGGGTAACTGCTGACCTGACACCTTCCCGGTGACACACTCACCACTGCT 960
Db 1599 CGCCCCATTTGCGCGGGTAACTGCTGACCTGACACCTTCCCGGTGACACACTCACCACTGCT 1658
QY 961 GCACCCCGGGGCTTCTGCGGCCCTCGGGCAGCGCTGCTGAGGAAGCCGAGGCCAGTGA 1020
Db 1659 GCACCCCGGGGCTTCTGCGGCCCTCGGGCAGCGCTGCTGAGGAAGCCGAGGCCAGTGA 1718
QY 1021 GCGTCTCAACGAGGCCAGCTCGCTGCACTCTCCGAGGTGCGGGCCACCGGCCG 1080
Db 1719 GCGTCTCAACGAGGCCAGCTCGCGCTGCACTCTCCGAGGTGCGGGCCACCGGCCG 1778
QY 1081 CCAGGGCTGGTGCCCGGCGACAGCTCATCCGAGCGCTGTTTACCAACAGGGCGGTGA 1140
Db 1779 CCAGGGCTGGTGCCCGGCGACAGCTCATCCGAGCGCTGTTTACCAACAGGGCGGTGA 1838
QY 1141 GAGCCTTGAGCACGGGGCCCTTGGCCCGGGGAGCAGCGCGCACCCGGCCCGGGCGGCA 1200
Db 1839 GAGCCTTGAGCACGGGGCCCTTGGCCCGGGGAGCAGCGCGCACCCGGCCCGGGCGGCA 1898
QY 1201 CTGTCTGAAGGGCGGTACATTTGGGGGGCGCTGCGCAGGCTGCGGAGCCACTGTATGA 1260
Db 1899 CTGTCTGAAGGGCGGTACATTTGGGGGGCGCTGCGCAGGCTGCGGAGCCACTGTATGA 1958
QY 1261 GCACCGGCTGCTGAGAGAGAGCGCGCAGGGAGGAGCGGCCACCCGGCCCGGGCGGCA 1320
Db 1959 GCACCGGCTGCTGAGAGAGAGCGCGCAGGGAGGAGCGGCCACCCCTCTGCGCAAGC 2018
QY 1321 CCCTCTCATTCGAGACTGCTCCCTCGGGTGTGCTGCTTGGCACCCACTTGGCCCTCGGCCA 1380
Db 2019 CCCTCTCATTCGAGACTGCTCCCTCGGGTGTGCTGCTTGGCACCCACTTGGCCCTCGGCCA 2078
QY 1381 CAGCCACTCCTCGAATCATGACTCTCGAGCAACCCCGGCCCTCTCTCGAGGCGCTGCGG 1440
Db 2079 CAGCCACTCCTCGAATCATGACTCTCGAGCAACCCCGGCCCTCTCTCGAGGCGCTGCGG 2138
QY 1441 TGAGGCACAGCGACTGCTTTCAGCCCTTCCGGGGGGCGCCCTATCAGGGACATGGGCA 1500
Db 2139 TGAGGCACAGCGACTGCTTTCAGCCCTTCCGGGGGGCGCCCTATCAGGGACATGGGCA 2198
QY 1501 CCTCAGGGCTCCAGAGCTTCCATCCACTGTGTGGCCACCCAGGCACTGCTCAGCCAGA 1560
Db 2199 CCTCAGGGCTCCAGAGCTTCCATCCACTGTGTGGCCACCCAGGCACTGCTCAGCCAGA 2258

QY 1561 GAGGCCATCCCGGACAGCCCTTGGGGGACAGCAGCCCTTTCTGCCACCCCAAGCAGGG 1620
Db 2259 GAGGCCATCCCGGACAGCCCTTGGGGGACAGCAGCCCTTTCTGCCACCCCAAGCAGGG 2318
QY 1621 TTCTGCCCCCCAGAGAGGGCTGACGCCCCCAACAGCAGTTGCCCATGCCCTCTCTGGTGC 1680
Db 2319 TTCTGCCCCCCAGAGAGGGCTGACGCCCCCAACAGCAGTTGCCCATGCCCTCTCTGGTGC 2378
QY 1681 CTTCCTCTCAGGATTTGCAAGAGGCCCCCTTAGTACCCTCAAGCCCCCTTTCTTGGGACA 1740
Db 2379 CTTCCTCTCAGGATTTGCAAGAGGCCCCCTTAGTACCCTCAAGCCCCCTTTCTTGGGACA 2438
QY 1741 GCCCAGGACCCCTGSCCTGSCCAAGCAAGCCCCCAATTTGGACTCTAAGATGGGGCC 1800
Db 2439 GCCCAGGACCCCTGSCCTGSCCAAGCAAGCCCCCAATTTGGACTCTAAGATGGGGCC 2498
QY 1801 TGGAGACATCTCTCTTCTGGGAGGCCAAACCCGCCCCCTGCAGTTCCCCCAGGGTCAGC 1860
Db 2499 TGGAGACATCTCTCTTCTGGGAGGCCAAACCCGCCCCCTGCAGTTCCCCCAGGGTCAGC 2558
QY 1861 CTCCAGGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 1920
Db 2559 CTCCAGGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 2618
QY 1921 AGAGCCTTGGGCCCTCCTGATGCGAGGSGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 1980
Db 2619 AGAGCCTTGGGCCCTCCTGATGCGAGGSGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 2678
QY 1981 CTCCACACCCACTTGGAGCGGCTCAGGAACAGGCGACCATCGCAAGTTCTCCCTGGG 2040
Db 2679 CTCCACACCCACTTGGAGCGGCTCAGGAACAGGTCAGCATCGCAAGTTCTCCCTGGG 2738
QY 2041 TGGTCCGGGGGCTACGACGGCTGCTGGCTGCTATGCGACCTTTGCTTGGTGGAGATGC 2100
Db 2739 TGGTCCGGGGGCTACGACGGCTGCTGGCTGCTATGCGACCTTTGCTTGGTGGAGATGC 2798
QY 2101 AGGGGCACTGCTGGGGCAGGGGCCCATGTGGGGCAGGATAGCTGGGCTGTGTCCAGTGC 2160
Db 2799 AGGGGCACTGCTGGGGCAGGGGCCCATGTGGGGCAGGATAGCTGGGCTGTGTCCAGTGC 2858
QY 2161 GAGGAGGAGGAGCAGGAGGAGCGCAGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2220
Db 2859 GAGGAGGAGGAGCAGGAGGAGCGCAGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2918
QY 2221 CAGGGCTGAGAGCCACTGCCCCAGTGCAGTGCAGAGGCTGTGCTTGGTGGAGAGGC 2280
Db 2919 CAGGGCTGAGAGCCACTGCCCCAGTGCAGTGCAGAGGCTGTGCTTGGTGGAGAGGC 2978
QY 2281 TCCACACAGGAGCTCTCCAGAGCCCAACCCATGGGAGGACATCGGGCAGGTCCTCCCTGGT 2340
Db 2979 TCCACACAGGAGCTCTCCAGAGCCCAACCCATGGGAGGACATCGGGCAGGTCCTCCCTGGT 3038
QY 2341 GCAGATCCGGGACCTGTGAGGTGATCGGAGGGCGCGCACAAATATCCCTGGACATTTTC 2400
Db 3039 GCAGATCCGGGACCTGTGAGGTGATCGGAGGGCGCGCACAAATATCCCTGGACATTTTC 3098
QY 2401 CGAGTGGAGCCCGGCTACTCAACTCTCAGACCTGTACAGATATCAAGTACTCCCATTT 2460
Db 3099 CGAGTGGAGCCCGGCTACTCAACTCTCAGACCTGTACAGATATCAAGTACTCCCATTT 3158
QY 2461 CGAGTTTATGATTTTCAAGAAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCCAT 2520
Db 3159 CGAGTTTATGATTTTCAAGAAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCCAT 3218
QY 2521 GCTCAGGAGGAGCTGGCGGAGTTCCCGAGGCCACGCTGGGCCCTGGCCAGGTGAATGGG 2580
Db 3219 GCTCAGGAGGAGCTGGCGGAGTTCCCGAGGCCACGCTGGGCCCTGGCCAGGTGAATGGG 3278
QY 2581 CCCCCCAGCAGGCTTGGAGATCACAGAGAGTTCAGAGGATGTGGACGCGCTGTCTGGCAGA 2640
Db 3279 CCCCCCAGCAGGCTTGGAGATCACAGAGAGTTCAGAGGATGTGGACGCGCTGTCTGGCAGA 3338

Qy 2641 GGCTGCCGTGGCAGGAGCGCAAGTGGTCTCGCCGTACGCGAGCCTCTTCCACTTCCC 2700
Dy 3339 GGCTGCCGTGGCAGGAGCGCAAGTGGTCTCGCCGTACGCGAGCCTCTTCCACTTCCC 3398
Qy 2701 TGGGAGGCACCTGCCGTGGATGAGCCTGACAGAGCTGGGGCTGGCGTGAGAGAGTGAAGGC 2760
Dy 3399 TGGGAGGCACCTGCCGTGGATGAGCCTGACAGAGCTGGGGCTGGCGTGAGAGAGTGAAGGC 3458
Qy 2761 CTCGCTGGAGCACATCTCCCGGATCTGAAGGGCAGCGCGAAGTCTGGAGAGAGGAGGG 2820
Dy 3459 CTCGCTGGAGCACATCTCCCGGATCTGAAGGGCAGCGCGAAGTCTGGAGAGAGGAGGG 3518
Qy 2821 GCGCCCGCAGGAAGCAGCGCTTGGTCTTCCGCTCTCAGTCTGAGAGAGTGGGA 2880
Dy 3519 GCGCCCGCAGGAAGCAGCGCTTGGTCTTCCGCTCTCAGTCTGAGAGAGTGGGA 3578
Qy 2881 CCGAGCCCGCAGATCTTCCCTAAGGGAGCTCTCAGATGAGACTGTGGTCTGGGCCAGTCAGT 2940
Dy 3579 CCGAGCCCGCAGATCTTCCCTAAGGGAGCTCTCAGATGAGACTGTGGTCTGGGCCAGTCAGT 3638
Qy 2941 GACACTGGCTGCAGGTGTGAGCCCGCAGCTGCCAGGCCACCTGGAGCAAGAGCG 3000
Dy 3639 GACACTGGCTGCAGGTGTGAGCCCGCAGCTGCCAGGCCACCTGGAGCAAGAGCG 3698
Qy 3001 A 3001
Dy 3699 A 3699
RESULT 8
AAD38865
ID AAD38865 standard; cdna; 5454 BP.
AC AAD38865;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human kinase (PKIN)-22.cdna.
XX
KW Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;
KW acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;
KW asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;
KW development; hepatitis; cardiovascular; hypertension; drug screening;
KW myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;
KW fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;
KW hypercholesterolaemia; obesity; gene therapy; cytostatic; anti-hiv;
KW neuroprotective; hepatotropic; hypotensive; cardiac; nephrotropic;
KW hyperlipidaemia; enzyme; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 373..5370
FT /tag= a
FT /product= "Human kinase (PKIN)-22"
XX
W0200233099-A2.
XX
PD 25-APR-2002.
XX
PF 20-OCT-2001; 2001WO-US047728.
XX
PR 20-OCT-2000; 2000US-0242410P.
PR 27-OCT-2000; 2000US-0244068P.
PR 03-NOV-2000; 2000US-0245708P.
PR 09-NOV-2000; 2000US-0247672P.
PR 16-NOV-2000; 2000US-0249565P.
PR 22-NOV-2000; 2000US-0252730P.
PR 01-DEC-2000; 2000US-0250807P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
PI Gururajan R, Baughn MR, Wallia NK, Elliott VS, Xu Y, Arvizu C;

PI Yao MG, Ramkumar J, Ding L, Tang YT, Hafalia AJA, Nguyen DB;
PI Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM, Lal PG;
PI Thangavelu K, Lu DAM, Borowsky ML, Thornton M, Swannaker A;
XX WPI; 2002-454603/48.
DR P-PSDB; AAE24151.
XX
PS New human kinase polypeptide, for diagnosing, preventing and treating
PS cancer, immune system disorders, growth and development disorders,
PS cardiovascular disorders and lipid disorders.
XX
XX Claim 5; Page 209-210; 210pp; English.
XX
CC The invention relates human kinases (PKIN) and their corresponding
CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,
CC treating and preventing cancer, an immune system disorder (e.g., acquired
CC immune deficiency syndrome (AIDS), Addison's disease, allergy, asthma,
CC atherosclerosis, multiple sclerosis, psoriasis), disorders affecting
CC growth and development (e.g., arteriosclerosis, cirrhosis, hepatitis),
CC cardiovascular disorder (e.g., hypertension, myocardial infarction,
CC Goodpasture's syndrome), and a lipid disorder (e.g., fatty liver,
CC Gaucher's disease, Niemann-Pick's disease, hypercholesterolaemia,
CC hyperlipidaemia, obesity), and for assessing the effects of exogenous
CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a
CC condition or a disease associated with the expression of PKIN in a
CC biological sample. A composition comprising PKIN or an agonist or
CC antagonist of PKIN is useful for treating a disease or condition
CC associated with decreased or increased expression of functional PKIN.
CC PKIN is useful in a number of drug screening techniques and to analyse
CC the proteome of a tissue or cell type. PKIN DNA is useful for creating
CC knockin humanised animals or transgenic animals to model human disease,
CC and in somatic or germline gene therapy. The present sequence is human
CC PKIN CDNA
XX
SQ Sequence 5454 BP; 1019 A; 1853 C; 1686 G; 896 T; 0 U; 0 Other;
Query Match 99.8%; Score 2996.2; DB 6; Length 5454;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 CCTACGAGCAGAACTCGGGCCAGGATACAGGAGGAGGAGATCTTGGCGCGCTGAG 60
Dy 963 CCTACGAGCAGAACTCGGGCCAGGATACAGGAGGAGGAGATCTTGGCGCGCTGAG 1022
Qy 61 CCACCCCGCTGTCAACGGGGCTGTGGACCAAGTTTGGAGCCGCAAGACCTCTCTCTCAT 120
Dy 1023 CCACCCCGCTGTCAACGGGGCTGTGGACCAAGTTTGGAGCCGCAAGACCTCTCTCTCAT 1082
Qy 121 CTGGAGCTGTCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAAGGCGGTGGTAC 180
Dy 1083 CTGGAGCTGTCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAAGGCGGTGGTAC 1142
Qy 181 GGAGCGGAGGTCAAGGTCTACATCCAGAGCTGTGGAGGGGTGCACTACCTGACAG 240
Dy 1143 GGAGCGGAGGTCAAGGTCTACATCCAGAGCTGTGGAGGGGTGCACTACCTGACAG 1202
Qy 241 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTTAACATCTCTGATGGTGCACTCTGCCCG 300
Dy 1203 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTTAACATCTCTGATGGTGCACTCTGCCCG 1262
Qy 301 GGAAGACATTAATAATCTGGACATTTGGCTTTGGCCAGAAACATCAACCCAGAGAGTGA 360
Dy 1263 GGAAGACATTAATAATCTGGACATTTGGCTTTGGCCAGAAACATCAACCCAGAGAGTGA 1322
Qy 361 GTTCAGCGAGTACGGCTCCCTGAGTTCTGCTCCCGAGATCATCCAGCAGAAACCTCTGT 420
Dy 1323 GTTCAGCGAGTACGGCTCCCTGAGTTCTGCTCCCGAGATCATCCAGCAGAAACCTCTGT 1382
Qy 421 GAGCGAAGCCTCCGACATTTGGGCCATGGGTGTCTCTCTTACCTCAGCCTGACCTGCTC 480
Dy 1383 GAGCGAAGCCTCCGACATTTGGGCCATGGGTGTCTCTCTTACCTCAGCCTGACCTGCTC 1442

QY 481 ATCCCAATTTGCGGCGAGAGTGACCGTGCACACCTCTGAAAGCTCCTGGAGGGCGCGT 540
Db 1443 ATCCCAATTTGCGGCGAGAGTGACCGTGCACACCTCTGAAAGCTCCTGGAGGGCGCGT 1502
QY 541 GTCATGAGCAGCCCATGGCTGCCACCTCAGCGAGAGCGCAAGACTTCAACAAGGC 600
Db 1503 GTCATGAGCAGCCCATGGCTGCCACCTCAGCGAGAGCGCAAGACTTCAACAAGGC 1562
QY 601 TAGCTGTCAGAGAGCCCTCAGGCCGCGCTAGTGGGCCCAAGTGCCTCTCCACCCCTG 660
Db 1563 TAGCTGTCAGAGAGCCCTCAGGCCGCGCTAGTGGGCCCAAGTGCCTCTCCACCCCTG 1622
QY 661 GTTCTGAAATCGATGCTGCGGAGGAGGCCCACTTCATCAACAACAAGCAGCTCAAGTT 720
Db 1623 GTTCTGAAATCGATGCTGCGGAGGAGGCCCACTTCATCAACAACAAGCAGCTCAAGTT 1682
QY 721 CCTCTGGGCCGAGTGCCTGACGCTTCCAGCTTCCCTGATGAGCTACAAAGTCCATCTGGTGAT 1742
Db 1683 CCTCTGGGCCGAGTGCCTGACGCTTCCAGCTTCCCTGATGAGCTACAAAGTCCATCTGGTGAT 1742
QY 781 GCGCTCCATTCCTGAGCTGCTGCGGGGCCCAACCCGACAGCCCTCCTCGGCGTAGCCCG 840
Db 1743 GCGCTCCATTCCTGAGCTGCTGCGGGGCCCAACCCGACAGCCCTCCTCGGCGTAGCCCG 1802
QY 841 GCACCTCTGACGAGACACTGGTGGCTCCTCGATTCCTCTCCTCTGCAACAAGAGCT 900
Db 1803 GCACCTCTGACGAGACACTGGTGGCTCCTCGATTCCTCTCCTCTGCAACAAGAGCT 1862
QY 901 CGCCCAATTTGCGGGCTTAAGTCACTGCGACCTCCCGGTGACACACTCACCACCTGCT 960
Db 1863 CGCCCAATTTGCGGGCTTAAGTCACTGCGACCTCCCGGTGACACACTCACCACCTGCT 1922
QY 961 GCAACCCCGGGCTTCTGCGGCCCTCGGCCAGCTGCTGAGGAAGCCGAGGCCAGTGA 1020
Db 1923 GCAACCCCGGGCTTCTGCGGCCCTCGGCCAGCTGCTGAGGAAGCCGAGGCCAGTGA 1982
QY 1021 GCGCTCCACCGAGGCCCGCTCGCTGCACTCCCGAGGCTGACACTCAACACCTGCT 1080
Db 1983 GCGCTCCACCGAGGCCCGCTCGCTGCACTCCCGAGGCTGACACTCAACACCTGCT 2042
QY 1081 CCAGGGCTGCTGCCCGGCAACAGCTCATCCGACGCTGTTCTACACACAGGCGGTGA 1140
Db 2043 CCAGGGCTGCTGCCCGGCAACAGCTCATCCGACGCTGTTCTACACACAGGCGGTGA 2102
QY 1141 GAGCCCTGAGCAGCGGGCTTGGCCCGGAGGAGCAGCGGCAACCGCCCGGGCGGCA 1200
Db 2103 GAGCCCTGAGCAGCGGGCTTGGCCCGGAGGAGCAGCGGCAACCGCCCGGGCGGCA 2162
QY 1201 CCTGCTGAAGGGCGGTACATTGGGGGGCGCTGCCAGGGCTGCGGAGCCTGATGGA 1260
Db 2163 CCTGCTGAAGGGCGGTACATTGGGGGGCGCTGCCAGGGCTGCGGAGCCTGATGGA 2222
QY 1261 GCAACCGCTGCTGAGGAGGAGCGCGCAGGAGGAGCAGGCCAACCTCTCTGGCCAAAGC 1320
Db 2223 GCAACCGCTGCTGAGGAGGAGCGCGCAGGAGGAGCAGGCCAACCTCTCTGGCCAAAGC 2282
QY 1321 CCCCTCAATTCGAGACTGCCCTCGGCTGCTGCTGAGCAGCCCACTTGGCCCTGGCCA 1380
Db 2283 CCCCTCAATTCGAGACTGCCCTCGGCTGCTGCTGAGCAGCCCACTTGGCCCTGGCCA 2342
QY 1381 CAGCCTCTCCCTGGAACATGACTCTCCGAGCAGCCCGCCGCTCTCTCGAGGCGCTGCGG 1440
Db 2343 CAGCCTCTCCCTGGAACATGACTCTCCGAGCAGCCCGCCGCTCTCTCGAGGCGCTGCGG 2402
QY 1441 TGAGGACACAGCATGCTCTTACGCCCTCCGGGGGGCCCTTATCAGGGACATGCGGCA 1500
Db 2403 TGAGGACACAGCATGCTCTTACGCCCTCCGGGGGGCCCTTATCAGGGACATGCGGCA 2462
QY 1501 CCTCAGGGCTCCAGCAGCTTCCATCCACTGCTGGCCACCCAGGCACTGCTCAGCCAGA 1560
Db 2463 CCTCAGGGCTCCAGCAGCTTCCATCCACTGCTGGCCACCCAGGCACTGCTCAGCCAGA 2522
QY 1561 GAGGCCATCCCGGCAAGCCCTTGGGGGCGAGCCAGCCCTTTCTGCCACCCCAAGCAGG 1620

Db 2523 GAGGCCATCCCGGACAGCCCTTGGGGGCGAGCCAGCCCTTTCTGCCACCCCAAGCAGG 2582
QY 1621 TTCTGCCCCCGAGGAGGGCTGCAGCCCCCAAGAGAGTTGCCCATTCCTCTGGCTC 1680
Db 2583 TTCTGCCCCCGAGGAGGGCTGCAGCCCCCAAGAGAGTTGCCCATTCCTCTGGCTC 2642
QY 1681 CTTCCTCTCAGGATCTTGCAGAGGGCCCTTAGTACCTCAAGCCCTCTTCTGGGACA 1740
Db 2643 CTTCCTCTCAGGATCTTGCAGAGGGCCCTTAGTACCTCAAGCCCTCTTCTGGGACA 2702
QY 1741 GCGCCAGGACCCCTCTGCCCTGCAAGAGCAAGCCCACTAGACTCTAAGATGGGGCC 1800
Db 2703 GCGCCAGGACCCCTCTGCCCTGCAAGAGCAAGCCCACTAGACTCTAAGATGGGGCC 2762
QY 1801 TGAGAGATCTCTCTCTGGAGGCAAAACCCGCGCCCTCGAGTTCCCAAGGGTCAGC 1860
Db 2763 TGAGAGATCTCTCTCTGGAGGCAAAACCCGCGCCCTCGAGTTCCCAAGGGTCAGC 2822
QY 1861 CTCCAGGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGCTGGGCTCCTCCAGTGGGCAC 1920
Db 2823 CTCCAGGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGCTGGGCTCCTCCAGTGGGCAC 2882
QY 1921 AGAGCCTTGGCCCTCTCGATCGGAGGGCTGGAACCCAGAGGGCTGAGGATCTGTCGA 1980
Db 2883 AGAGCCTTGGCCCTCTCGATCGGAGGGCTGGAACCCAGAGGGCTGAGGATCTGTCGA 2942
QY 1981 CTCACACCCACTTTCAGAGGGCTCAGGAACAGGCGACCATCGCAAGTTCTTCCCTGGG 2040
Db 2943 CTCACACCCACTTTCAGAGGGCTCAGGAACAGGCGACCATCGCAAGTTCTTCCCTGGG 3002
QY 2041 TGGTCCCGGGGCTACGAGGGCTGCTGCTATGCACTTTCCTTTGGTGGAGATGC 2100
Db 3003 TGGTCCCGGGGCTACGAGGGCTGCTGCTATGCACTTTCCTTTGGTGGAGATGC 3062
QY 2101 AGGGGCACTGCTGGGGCAGGGGCCCATGTTGGGCCAGGATAGCTTGGCTGTGTCAGTTC 2160
Db 3063 AGGGGCACTGCTGGGGCAGGGGCCCATGTTGGGCCAGGATAGCTTGGCTGTGTCAGTTC 3122
QY 2161 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2220
Db 3123 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3182
QY 2221 CAGGGCTGAGAGCCCACTGCCCCAGTCACTGCAAGGCTGCTGCTGAGGTCGGCAGGCG 2280
Db 3183 CAGGGCTGAGAGCCCACTGCCCCAGTCACTGCAAGGCTGCTGCTGAGGTCGGCAGGCG 3242
QY 2281 TCCCAACAGGAGCTCTCAGAGCCCAACCCATGGAGGACATCGGGCAGGTCCTCTGGT 2340
Db 3243 TCCCAACAGGAGCTCTCAGAGCCCAACCCATGGAGGACATCGGGCAGGTCCTCTGGT 3302
QY 2341 GCAGATCCGGGACCTCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2400
Db 3303 GCAGATCCGGGACCTCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2460
QY 2401 CGAGTGGAGCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db 3363 CGAGTGGAGCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 3422
QY 2461 CGAGTTTATGATCTTTCAGGAAGTCCCAAGTCCGCTCAGCCAGAGCCGCTCCCCCAT 2520
Db 3423 CGAGTTTATGATCTTTCAGGAAGTCCCAAGTCCGCTCAGCCAGAGCCGCTCCCCCAT 3482
QY 2521 GGCTGAGGAGGAGTGGCGGAGTTCCCGAGCCCAAGTGGCCCTGGCCAGGTCGACTGGG 2580
Db 3483 GGCTGAGGAGGAGTGGCGGAGTTCCCGAGCCCAAGTGGCCCTGGCCAGGTCGACTGGG 3542
QY 2581 CCCCCACAGCGCTGGAGATCAAGAGGAGTCAAGAGGATGAGAGCGCTGCTGGCAGA 2640
Db 3543 CCCCCACAGCGCTGGAGATCAAGAGGAGTCAAGAGGAGTCAAGAGGAGTCAAGAGGAGT 3602
QY 2641 GGCTGCGTGGGAGGAGGAGGAGTGGTCTCTCGCGCTCAGCGAGCCCTCTTCCACTTCCC 2700

Db	725	CGCCCAAGGGCTGCGTGCGCCCGGCAACAGCGTCA	TCCGAGGCGCTGTCTTCA	CACCAAGGCGG	784
Qy	1137	GTGAGAGCGCTTGAGCA	CGGGGCGCTTGGCCCGGGGAGCAGCGGCA	CGCCGCGCGCGCGCGC	1196
Db	785	GTGAGAGCGCTTGAGCA	CGGGGCGCTTGGCCCGGGGAGCAGCGGCA	CGCCGCGCGCGCGCGC	844
Qy	1197	GGACCTGCTGAGGGCGGGTACAT	TGCGGGGGCGCTGCCAGGCGTGGCGAGCACTGA	1256	
Db	845	GGACCTGCTGAAGGCGGGTACAT	TGCGGGGGCGCTGCCAGGCGTGGCGAGCACTGA	904	
Qy	1257	TGGAGCACGGGTGTGAGGAGGAGGCGCGC	CAGGAGGAGGAGGCGCA	CCCTCTCTGCGCA	1316
Db	905	TGGAGCACGGGTGTGAGGAGGAGGCGCGC	CAGGAGGAGGAGGCGCA	CCCTCTCTGCGCA	964
Qy	1317	AAGCCCCCTCATTCAGAGACTGC	CCCTCGGCTTGCCTCTGCGACACCACTTGGCCCTG	1376	
Db	965	AAGCCCCCTCATTCAGAGACTGC	CCCTCGGCTTGCCTCTGCGACACCACTTGGCCCTG	1024	
Qy	1377	GCCACAGCCA	CTCCCTGGAACATGACTCTCCGAGCAC	CCCCCGCCCTCTCTCGAGAGCGCT	1436
Db	1025	GCCACAGCCA	CTCCCTGGAACATGACTCTCCGAGCAC	CCCCCGCCCTCTCTCGAGAGCGCT	1084
Qy	1437	GCGGTGAGGCACAGGCACTG	CGCCCTTACGCCCTCTGGGGGGGCGCCCTAT	CAAGGACATGG	1496
Db	1085	GCGGTGAGGCACAGGCACTG	CGCCCTTACGCCCTCTGGGGGGGCGCCCTAT	CAGGACATGG	1144
Qy	1497	GGCACCTCTAGGGTCTCAAGCAGCTT	CCATCCATCTGTGGGCCACCCAGGACATGCTCAGC	1556	
Db	1145	GGCACCTCTAGGGTCTCAAGCAGCTT	CCATCCATCTGTGGGCCACCCAGGACATGCTCAGC	1204	
Qy	1557	CAGAGAGGCCATCCCGGACAGCCCTT	TGGGGGACAGCAGCCCTTTCTGCCACCCCAAGC	1616	
Db	1205	CAGAGAGGCCATCCCGGACAGCCCTT	TGGGGGACAGCAGCCCTTTCTGCCACCCCAAGC	1264	
Qy	1617	AGGGTCTGCCCCCAGAGGGCTG	CAGCCCCCACCAGCATGTTGCCCATGCTCCTG	1676	
Db	1265	AGGGTCTGCCCCCAGAGGGCTG	CAGCCCCCACCAGCATGTTGCCCATGCTCCTG	1324	
Qy	1677	GCTCCTTCCCTCCAGGATCTTTG	CAAGAGGCGCCCTTAGTACCTCAAGCCCTCTCTTGG	1736	
Db	1325	GCTCCTTCCCTCCAGGATCTTTG	CAAGAGGCGCCCTTAGTACCTCAAGCCCTCTCTTGG	1384	
Qy	1737	GACAGCCCCAGGCA	CCCCCTGCCAAAGCAAGCCCCCATTTGGA	CTCTAAGATGG	1796
Db	1385	GACAGCCCCAGGCA	CCCCCTGCCAAAGCAAGCCCCCATTTGGA	CTCTAAGATGG	1444
Qy	1797	GGCTTGAGACATCTCTTCT	TGGGAGGCGCAAAACCGGCGCCCTG	CAGTTCGCCAGGGT	1856
Db	1445	GGCTTGAGACATCTCTCTTCT	TGGGAGGCGCAAAACCGGCGCCCTG	CAGTTCGCCAGGGT	1504
Qy	1857	CAGCCTCCAGGCGAGCTCTT	CCCAAGTAGCTCCCTCAGGGTGGGCTCTCTCC	CCAGGTGG	1916
Db	1505	CAGCCTCCAGGCGAGCTCTT	CCCAAGTAGCTCCCTCAGGGTGGGCTCTCTCC	CCAGGTGG	1564
Qy	1917	GCA	CAGAGCCTGCGCCCTCGATGCGGAGGGCTG	GAACCCAGGAGGCTGAGGATCTGT	1976
Db	1565	GCA	CAGAGCCTGCGCCCTCGATGCGGAGGGCTG	GAACCCAGGAGGCTGAGGATCTGT	1624
Qy	1977	CCGACTCCACACCA	CTTGCAGGGCGCTCAGGAACAGGGCA	CCATGCGCAAGTTCTCC	2036
Db	1625	CCGACTCCACACCA	CTTGCAGGGCGCTCAGGAACAGGGCA	CCATGCGCAAGTTCTCC	1684
Qy	2037	TGGGTGTCGCGGGGCTAC	GAGGGCTAGCTGCTATGGCACTTGTGGTGGAG		2096
Db	1685	TGGGTGTCGCGGGGCTAC	GAGGGCTAGCTGCTATGGCACTTGTGGTGGAG		1744
Qy	2097	ATC	CAGGGGCGATGCTGGGGCAGGGGCCATGTGGGCG	CAGGATAGCTGGGCTGTCTCC	2156
Db	1745	ATC	CAGGGGCGATGCTGGGGCAGGGGCCATGTGGGCG	CAGGATAGCTGGGCTGTCTCC	1804
Qy	2157	AGTCGAGGAGGAGGAC	GAGGAGGCGCTCAGTCCAGTCCGAGGAGCAGCAGG		2216
Db	1805	AGTCGAGGAGGAGGAC	GAGGAGGCGCTCAGTCCAGTCCGAGGAGCAGCAGG		1864

QY	2217	AGGCCAGGGCTGAGAGCCCACTGCCCCAGGTCAGTGCAAGGCCCTGTGCGCTGAGGTCGGCA	2276
Db	1865	AGGCCAGGGCTGAGAGCCCACTGCCCCAGGTCAGTGCAAGGCCCTGTGCGCTGAGGTCGGCA	1924
QY	2277	GGGCTCCCAACCAAGAGCTCTCCAGAGCCCAACCCATGGGAGGACATCGGCGCAGGTCCTCC	2336
Db	1925	GGGCTCCCAACCAAGAGCTCTCCAGAGCCCAACCCATGGGAGGACATCGGCGCAGGTCCTCC	1984
QY	2337	TGGTGCAATGCCGGGAACCTGTCAAGTGATGCGGAGGCGGCCGACAAATATCCCTGGACA	2396
Db	1985	TGGTGCAATGCCGGGAACCTGTCAAGTGATGCGGAGGCGGCCGACAAATATCCCTGGACA	2044
QY	2397	TTTCCGAGGTGGACCCCGCCTACTCAACTCTCAGACCTGTACGATATCAAGTACCTCC	2456
Db	2045	TTTCCGAGGTGGACCCCGCCTACTCAACTCTCAGACCTGTACGATATCAAGTACCTCC	2104
QY	2457	CATTTCGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGTCAGCCAGAGCGCCCTCC	2516
Db	2105	CATTTCGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGTCAGCCAGAGCGCCCTCC	2164
QY	2517	CCATGTGCTGAGGAGAGCTGGCCGAGTTCCCGAGGCCACGTCGGCCCTGGCCAGGTGAAC	2576
Db	2165	CCATGTGCTGAGGAGAGCTGGCCGAGTTCCCGAGGCCACGTCGGCCCTGGCCAGGTGAAC	2224
QY	2577	TGGGCCCCCAGCAGAGCCCTGGAGATCACAGAGGAGTCAGAGATGTGGACGCGCTGCTGG	2636
Db	2225	TGGGCCCCCAGCAGAGCCCTGGAGATCACAGAGGAGTCAGAGATGTGGACGCGCTGCTGG	2284
QY	2637	CAGAGGCTGCGGTGGGCAAGAACGCAAGTGGTTCCTCGCCGTCAGCAGCCCTTTCCACT	2696
Db	2285	CAGAGGCTGCGGTGGGCAAGAACGCAAGTGGTTCCTCGCCGTCAGCAGCCCTTTCCACT	2344
QY	2697	TCCCTGGGAGGCACTTGCCTGGATGAGCGCTGCAGAGCTGGGGCTGCGTGAGAGAGTGA	2756
Db	2345	TCCCTGGGAGGCACTTGCCTGGATGAGCGCTGCAGAGCTGGGGCTGCGTGAGAGAGTGA	2404
QY	2757	AGGCTCCGTGGAGCACATCTCCCGGATCCTGAAGGGCAGGCCGAGGTCGGAGAAGG	2816
Db	2405	AGGCTCCGTGGAGCACATCTCCCGGATCCTGAAGGGCAGGCCGAGGTCGGAGAAGG	2464
QY	2817	AGGGGCCCCCAGGAAGAACAGGCTTGCTTCCTTCGGGCTCTCAGGTCCTGAAGAGCT	2876
Db	2465	AGGGGCCCCCAGGAAGAACAGGCTTGCTTCCTTCGGGCTCTCAGGTCCTGAAGAGCT	2524
QY	2877	GGGACCGAGCGCCGACATTCCTAAGGAGCTCTCAGATGAGACTGTGTCTCTGGGCCAGT	2936
Db	2525	GGGACCGAGCGCCGACATTCCTAAGGAGCTCTCAGATGAGACTGTGTCTCTGGGCCAGT	2584
QY	2937	CAGTGACACTGGCCTGCCAGGTGTGAGCCAGCCAGCTGCCAGGCCACCTGGAGCAAG	2996
Db	2585	CAGTGACACTGGCCTGCCAGGTGTGAGCCAGCCAGCTGCCAGGCCACCTGGAGCAAG	2644
QY	2997	ACGGA 3001	
Db	2645	ACGGA 2649	
RESULT 10			
AAD26467			
ID	AAD26467	standard; cDNA; 3225 BP.	
XX	AC	AAD26467;	
XX	AC	AAD26467;	
XX	DT	26-MAR-2002 (first entry)	
XX	DE	Human kinase PKIN-20 cDNA.	
XX	KW	Human; kinase; PKIN-20; cancer; leukaemia; adenocarcinoma; osteoporosis;	
XX	KW	immune disorder; atherosclerosis; Crohn's disease; Hodgkin's disease;	
XX	KW	Acquired Immune Deficiency Syndrome; AIDS; Addison's disease; anaemia;	
XX	KW	allergy; asthma; adult respiratory distress syndrome; multiple sclerosis;	
XX	KW	autoimmune thyroiditis; bronchitis; diabetes mellitus; osteoarthritis;	

KW	Good pasture's syndrome; Graves' disease; pancreatitis; psoriasis;
KW	rheumatoid arthritis; ulcerative colitis; cirrhosis; Cushing's syndrome;
KW	hemiparesis; hypothyroidism; cerebral palsy; cataract; angina pectoris;
KW	cardiovascular disease; hypertension; vasculitis; myocarditis; obesity;
KW	congestive heart failure; ischaemic heart disease; lung tumour; gout;
KW	fatty liver; Niemann-Pick's disease; gene therapy; ss.
OS	Homo sapiens.
XX	
XX	Location/Qualifiers
FT	526..3141
FT	/ *tag= a
FT	/ product= "Human PKIN-20 protein"
XX	
PN	WO200196547-A2.
XX	
PD	20-DEC-2001.
XX	
PP	14-JUN-2001; 2001WO-US019444.
XX	
PR	15-JUN-2000; 2000US-0212073P.
PR	23-JUN-2000; 2000US-0213467P.
PR	30-JUN-2000; 2000US-0215651P.
PR	07-JUL-2000; 2000US-0216605P.
PR	13-JUL-2000; 2000US-0218372P.
PR	25-AUG-2000; 2000US-0228056P.
XX	
XX	(INCY-) INCYTE GENOMICS INC.
PA	
PI	Yue H, Lal P, Bandman O, Borowsky ML, Au-Young J, Lu Y;
PI	Gandhi AR, Tribouley CM, Wallia NK, Yao MG, Lu DAM, Greenwald SR;
PI	Ramkumar J, Griffin JA, Kearney L, Burford N, Nguyen DB, Tang YT;
PI	Baughn MR, He A, Thornton M, Hafalia A, Patterson C, Gururajan R;
PI	Lo TP, Khan F, Recipon SA, Azimzai Y, Policky JL, Ding L;
PI	Grether M, Elliott VS, Thangavelu K, Batra S, Ison CH,
XX	
DR	WPI; 2002-090207/12.
DR	P-PSDB; AE16274.
XX	
PT	New polypeptides, useful for diagnosing, treating or preventing disorders
PT	of growth and development, cardiovascular and lipid, and diseases such as
PT	cancer, comprise human kinase polypeptides.
XX	
XX	Claim 5; Page 191; 197pp; English.
XX	
CC	The invention relates to human kinase PKIN proteins and their
CC	corresponding cDNAs. A composition containing PKIN agonist is useful for
CC	treating a disease or condition associated with decreased expression of
CC	PKIN and a composition comprising PKIN antagonist is useful for treating
CC	a disease or condition associated with overexpression of PKIN. The
CC	disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma,
CC	myeloma, sarcoma, teratocarcinoma, Hodgkin's disease); immune disorder
CC	(Acquired Immune Deficiency Syndrome (AIDS), asthma, Addison's disease,
CC	atherosclerosis, anaemia, allergies, adult respiratory distress syndrome,
CC	autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes
CC	mellitus, multiple sclerosis, Good pasture's syndrome, Graves' disease,
CC	osteoarthritis, osteoporosis, pancreatitis, psoriasis, Reiter's syndrome,
CC	rheumatoid arthritis, Sjogren's syndrome, uveitis, ulcerative colitis,
CC	bacterial, parasitic, fungal, viral, protozoal and helminthic infections)
CC	Growth and development disorders (arteriosclerosis, cirrhosis, hepatitis,
CC	Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts); cardio
CC	vascular disease (arteriovenous fistula, hypertension, vasculitis,
CC	aneurysms, congestive heart failure, angina pectoris, myocarditis,
CC	ischaemic heart disease, chronic bronchitis, lung tumours); lipid
CC	disorder (fatty liver, Fabry's disease, Niemann-Pick's disease,
CC	hypocholesterolaemia, obesity). PKIN DNA is useful for assessing toxicity
CC	of a test compound and in gene therapy. The present sequence is human
CC	PKIN-20 CDNA
XX	
SQ	Sequence 3225 BP; 596 A; 1104 C; 1017 G; 508 T; 0 U; 0 Other;

Query Match Score 1706.8; DB 6; Length 3225;
Best Local Similarity 99.8%; Pred No. 6.5e-297;

QY 2360 GGTGATGCGAGGCGGCGGACACAATATCCTCTGGACATTTCCGAGGTGGACCCCGCCTAC 2419
Db |||||
QY 1093 GGTGATGCGAGGCGGCGGACACAATATCCTCTGGACATTTCCGAGGTGGACCCCGCCTAC 1152
Db |||||
QY 2420 CTCACCTCTCAGACCTGTACGATATCAAGTACCTCCCATTCGAGTTTATGATCTTCAGG 2479
Db |||||
QY 1153 CTCACCTCTCAGACCTGTACGATATCAAGTACCTCCCATTCGAGTTTATGATCTTCAGG 1212
Db |||||
QY 2480 AAGTCCCAAGTCCGCTCAGCCAGAGCGGCCCTCCCCCATGGCTGAGGAGGAGCTGSCC 2539
Db |||||
QY 1213 AAGTCCCAAGTCCGCTCAGCCAGAGCGGCCCTCCCCCATGGCTGAGGAGGAGCTGSCC 1272
Db |||||
QY 2540 GAGTTCCCGAGCCCACTGCGCCCTGGCCAGGTGAATCGGCCCCCAGCGAGGCTTGGAG 2599
Db |||||
QY 1273 GAGTTCCCGAGCCCACTGCGCCCTGGCCAGGTGAATCGGCCCCCAGCGAGGCTTGGAG 1332
Db |||||
QY 2600 ATCAGAGGAGTACAGAGGATGGAGCGCTGTGGCAGAGGCTGCCGTGGGAGGAG 2659
Db |||||
QY 1333 ATCAGAGGAGTACAGAGGATGGAGCGCTGTGGCAGAGGCTGCCGTGGGAGGAG 1392
Db |||||
QY 2660 CGAAGTGGTCTCGCCCTCAGCAGCGCTCTTCCACTTCCCTGGGAGGCACTGCCGCTG 2719
Db |||||
QY 1393 CGAAGTGGTCTCGCCCTCAGCAGCGCTCTTCCACTTCCCTGGGAGGCACTGCCGCTG 1452
Db |||||
QY 2720 GATGAGCTTCAGAGCTGGGCTCGTGAGAGAGTGAAAGGCTCCGTGGAGCACATCTCC 2779
Db |||||
QY 1453 GACGAGCTTCAGAGCTGGGCTCGTGAGAGAGTGAAAGGCTCCGTGGAGCACATCTCC 1512
Db |||||
QY 2780 CGGATCTCTGAAGGCGAGCGCGGAGGCTCTGAGAGGAGGCGGCCCCCAGGAGAGCCCA 2839
Db |||||
QY 1513 CGGATCTCTGAAGGCGAGCGCGGAGGCTCTGAGAGGAGGCGGCCCCCAGGAGAGCCCA 1572
Db |||||
QY 2840 GGCCTTGTCTCTTCCGCTCTCAGGCTCTGAAGAGCTGGGAGCGGCGGACATTCCTA 2899
Db |||||
QY 1573 GGCCTTGTCTCTTCCGCTCTCAGGCTCTGAAGAGCTGGGAGCGGCGGACATTCCTA 1632
Db |||||
QY 2900 AGGAGCTCTCAGATGAGACTGTGGCTCGTGCCGAGTCACTGACACTGTGCGCTGCCAGGTG 2959
Db |||||
QY 1633 AGGAGCTCTCAGATGAGACTGTGGCTCGTGCCGAGTCACTGACACTGTGCGCTGCCAGGTG 1692
Db |||||
QY 2960 TCAGCCAGCGAGCTGCCAGGCGCACCTGGAGGCAAGACCGGA 3001
Db |||||
QY 1693 TCAGCCAGCGAGCTGCCAGGCGCACCTGGAGGCAAGACCGGA 1734
Db |||||

RESULT 11
AAF44662
ID AAF44662 standard; cdna; 7710 BP.
XX
XX AAF44662;
AC
XX
XX 27-MAR-2001 (first entry)
DT
DE Novel protein kinase cdna, SEQ ID NO: 42.
XX
XX Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
KW dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; SS.
XX
XX Homo sapiens.
OS
XX
XX WO200073469-A2.
PN
XX
XX 07-DEC-2000.
PD
XX
XX 26-MAY-2000; 2000WO-US014842.
PF
XX
XX 28-MAY-1999; 99US-0136503P.
PR
XX
XX

(SUGE-) SUGEN INC.

Plowman GD, Martinez R, Whyte D, Sudersanam S;

WPI; 2001-032161/04.

P-PSDB; AAB65635.

Nucleic acids encoding kinase polypeptides, useful for diagnosing and treating immune-related diseases and disorders, cardiovascular disease, neurodegenerative diseases and/or cancers.

Example 1; Fig 2; 310pp; English.

The present sequence encodes a novel protein kinase. The nucleic acids and the protein kinases they encode may be used in the treatment and diagnosis of diseases associated with inappropriate kinase expression such as immune-related diseases and disorders, cardiovascular disease, neurodegenerative diseases and/or cancers. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays. The kinase polypeptides may be used as antigens in the production of antibodies of kinase expression and activity. Anti-kinase antibodies of kinases antagonists may also be used to down regulate kinase expression and activity. Diseases related to kinase expression and activity include rheumatoid arthritis, atherosclerosis, autoimmune disorders, complications of organ transplantation, myocardial infarction, immune disorders, cardiomyopathies, strokes, renal failure, oxidative-stress related disorders, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and reproductive disorders

Sequence 7710 BP; 1446 A; 2534 C; 2447 G; 1283 T; 0 U; 0 Other;

Query Match 7.4%; Score 223.2; DB 4; Length 7710;
Best Local Similarity 59.0%; Pred. No. 1.2e-30;
Matches 466; Conservative 0; Mismatches 303; Indels 21; Gaps 4;

QY 33 GGGAGCGAGACATCTCTGGCGGCTGAGCCACCCGCTGGTGCACGGGGCTGCTGGACAGT 92
Db |||||
QY 1988 GGGAGCGCGGCTGCTGGCCAGGCTCCAGCAGCAGTGTCTCTACTTCCATGAGGCCT 2047
Db |||||
QY 93 TTGAGACCCGCAAGACCCCTCATCTCTGGAGCTGTGCTCATCCGAGGAGCTGCTGG 152
Db |||||
QY 2048 TCGAGAGGCGCGGGAGCTGGTCAATTGTTCAGGAGCTCTGCACA---GAGGAGCTGCTGG 2104
Db |||||
QY 153 ACCGCTGTACAGGAAGGCGGTGTGACGAGGCGGAGTCAAGGTCTACATCCAGCAGC 212
Db |||||
QY 2105 AGCGAATCCGAGGAACCCACCGTGTGTGATCTGAGATCCGGGCTATATGCGGCGAG 2164
Db |||||
QY 213 TGGTGGAGGGCTGCACTACCTGCACAGCCATGGCGTTCTCCACCTGGACATAAAGCCCT 272
Db |||||
QY 2165 TGCTAGAGGGAATACACTACCTGCACAGAGCCAGTGTGCACCTCGATGTCAAGCCTG 2224
Db |||||
QY 273 CTAACATCTGATGTGTCATCTCTG-----CGGGAAGACATTAATCTGCGACTTTG 326
Db |||||
QY 327 GCTTTGCCAGAAATCACCCAGCAGAGCTGCAGTTACAGCAGTACGCTCCCTGACT 386
Db |||||
QY 2285 GGAATGCCAGGAGCTGACTCCAGGAGAGCCCACTACTGCCAGTATGTCACACCTGAGT 2344
Db |||||
QY 387 TCGTCTCCCGGAGATCATCCAGCAGAAACCTGTGTAGCGGAAGCCTCCGACATTTGGGCCA 446
Db |||||
QY 2345 TTGTAGCACCCGAGATTGTCAATCAGAGCCCGTGTCTGGAGTCACTGACATCTGCGCTG 2404
Db |||||
QY 447 TGGGTGTCTCTCTACCTCAGCTGACTGCTCATCCCATTTTCCCGCGAGAGTGACC 506
Db |||||
QY 2405 TGGGTGTGTTGCTCTCTCTGTGTGACAGGAATCTCCCGCTTTGTTGGGGAATGACC 2464
Db |||||
QY 507 GTGCCACCCCTCTCTGAACGCTCTGGAGGGCGGCTCATGTGGAGCAGCCCATGGTGCCC 566
Db |||||
QY 2465 GGACAAACATTGATGAACATCCGAACTACAACGCTGGCCTTCGAGGAGACCAATTCCTGA 2524
Db |||||

XX	Human; kinase polypeptide; PKIN-18; gene therapy; Addison's disease;
KW	leukemia; immune disorder; lymphoma; melanoma; developmental disorder;
KW	acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hypertension;
KW	asthma; Crohn's disease; rheumatoid arthritis; bursitis; atherosclerosis;
KW	cirrhosis; hepatitis; psoriasis; Cushing's syndrome; cycostatic; cancer;
KW	cholestasis; cardiac; cardiovascular disorder; Niemann-Pick's disease;
KW	lipid disorder; fatty liver; Gaucher's disease; myocardial infarction;
KW	hypertension; transgenic animal; antiinflammatory; hepatotropic;
KW	hypotensive; anti-HIV; enzymes; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200208399-A2.
XX	
PD	31-JAN-2002.
XX	
PF	20-JUL-2001; 2001WO-US023092.
XX	
PR	21-JUL-2000; 2000US-0220038P.
PR	28-JUL-2000; 2000US-0222112P.
PR	04-AUG-2000; 2000US-0222831P.
PR	11-AUG-2000; 2000US-0224729P.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
PA	(THOR-) THORNTON M.
XX	
PI	Thornton M, Yue H, Khan FA, Gururajan R, Hafalia AJA, Walia NK;
PI	Patterson C, Ramkumar J, Gandhi AR, Policky JL, Baughn MR;
PI	Tribolety CM, Bandman O, Nguyen DB, Lu Y, Burford N, Lal P, Ding L;
PI	Yao MG, Elliott VS, Recipon SA, Kearney L, Lu DAM, Greenwald SR;
PI	Tang YT, Xu Y, Walsh RT, Gietzen KJ, Yang J, Hillman JL;
XX	
DR	WPI; 2002-206083/26.
XX	
PT	New human kinase polypeptide, useful in diagnosis, prevention and
PT	treatment of cancer, immune disorder, growth and developmental disorder,
PT	cardiovascular disorder and lipid disorder.
XX	
FS	Claim 5; Page 191-193; 196pp; English.
XX	
CC	The present invention relates to an isolated human kinase polypeptide
CC	(PKIN) or a biologically active/immunogenic fragment of PKIN. PKIN is
CC	useful for diagnosing, treating and preventing cancer (e.g., leukaemia,
CC	lymphoma, melanoma), an immune disorder (e.g., acquired immunodeficiency
CC	syndrome (AIDS), Addison's disease, allergy, anaemia, asthma, Crohn's
CC	disease, rheumatoid arthritis), a growth and developmental disorder (e.g.
CC	bursitis, cirrhosis, hepatitis, psoriasis, Cushing's syndrome), a
CC	cardiovascular disorder (e.g., atherosclerosis, hypertension, myocardial
CC	infarction), and a lipid disorder (e.g., fatty liver, cholestasis,
CC	Gaucher's disease, Niemann-Pick's disease). PKIN is useful in a number of
CC	drug screening techniques and to analyse the proteome of a tissue or cell
CC	type. PKIN is useful for creating knockin humanised animals or transgenic
CC	animals to model human diseases, in somatic or germline gene therapy, to
CC	generate a transcript image of a tissue or cell type for detecting
CC	differences in the chromosomal location due to translocation, inversion,
CC	etc., among normal, carrier or affected individuals, and as hybridisation
CC	probes for mapping naturally occurring genomic sequences. PKIN is useful
CC	in southern or northern analysis, dot blot or other membrane-based
CC	technologies, in PCR technologies, in dipstick, pin, microformat enzyme
CC	linked immunosorbent (ELISA)-like assays and in microarrays utilising
CC	fluids or tissues from patients to detect altered PKIN expression. The
CC	present sequence is human PKIN-18 cDNA. Note: This sequence is said to
CC	encode PKIN-18 referred as SEQ ID NO:18 (AAE19160). However this does not
CC	appear to be the case
XX	
SQ	Sequence 7789 BP; 1469 A; 2558 C; 2470 G; 1292 T; 0 U; 0 Other;
Query Match	7.4%; Score 223.2; DB 6; Length 7789;
Best Local Similarity	59.0%; Pred. No. 1.2e-30;
Matches	466; Conservative 0; Mismatches 303; Indels 21; Gaps 4;
QY	33 GSGAGCGAGACATCTGGCGCGCTGAGCCACCGCTGGTGTCACGGGGCTCTCGACCCAGT 92

XX PN WO2003076642-A2.
 XX FD 18-SEP-2003.
 XX XX 02-AUG-2002; 2002WO-US024459.
 XX PR 02-AUG-2001; 2001US-0309501P.
 XX PR 03-AUG-2001; 2001US-0310291P.
 XX PR 08-AUG-2001; 2001US-0310951P.
 XX PR 09-AUG-2001; 2001US-0311292P.
 XX PR 13-AUG-2001; 2001US-0311979P.
 XX PR 14-AUG-2001; 2001US-0312203P.
 XX PR 17-AUG-2001; 2001US-0313156P.
 XX PR 17-AUG-2001; 2001US-0313201P.
 XX PR 20-AUG-2001; 2001US-0313702P.
 XX PR 21-AUG-2001; 2001US-0314031P.
 XX PR 23-AUG-2001; 2001US-0314466P.
 XX PR 28-AUG-2001; 2001US-0315403P.
 XX PR 29-AUG-2001; 2001US-0315853P.
 XX PR 31-AUG-2001; 2001US-0316508P.
 XX PR 21-SEP-2001; 2001US-0323936P.
 XX PR 03-DEC-2001; 2001US-0338078P.
 XX PR 05-FEB-2002; 2002US-0334655P.
 XX PR 05-MAR-2002; 2002US-0361764P.
 XX PR 19-APR-2002; 2002US-0373825P.
 XX PR 15-MAY-2002; 2002US-0380971P.
 XX PR 15-MAY-2002; 2002US-0380980P.
 XX PR 16-MAY-2002; 2002US-0381039P.
 XX PR 28-MAY-2002; 2002US-0383761P.
 XX PR 29-MAY-2002; 2002US-0383887P.
 XX PR 01-AUG-2002; 2002US-00210130.
 XX PA (CURA-) CURAGEN CORP.
 XX XX Zetshusen BD, Patturajan M, Kekuda R, Miller CE, Rieger DK;
 XX PI Pena CE, Shimkets RA, Li L, Berghs C, Zhong M, Caeman SJ, Voss EZ;
 XX PI Boldog FL, Padigaru M, Smithson G, Shenoy SG, Ji W, Gorman L;
 XX PI Vernet CAM, Leite MW, Guo X, Anderson DW, Spytek KA, Gerlach VL;
 XX PI Burgess CB, Khrantsov NV, Ort T, Ellerman K, Rastelli L, Agee ML;
 XX PI Chaudhuri A, Chant JS, Dipippo VA, Edinger SR, Eisen A, Gangolli EA;
 XX PI Giot L, Ooi CE, Rothenberg ME, Spaderna SK, Hjalt T, Liu X;
 XX PI Taupier RJ, Catterton E;
 XX XX WPI; 2003-779062/73.
 XX DR P-PSDB; ADE47676.
 XX XX
 XX PT New NOVX polypeptides and nucleic acids, useful for preventing or
 XX PT treating NOVX-associated disorders, e.g. cancer, diabetes,
 XX PT atherosclerosis, asthma or AIDS, and in chromosome mapping, tissue typing
 XX PT or pharmacogenomics.
 XX XX Claim 20; SEQ ID NO 37; 562pp; English.
 XX PS
 XX XX The invention relates to a novel (NOVX) human polypeptide. A polypeptide
 XX CC of the invention has cardiac, antiarteriosclerotic, hypotensive,
 XX CC immunosuppressive, dermatological, anorectic, cytostatic, antidiabetic,
 XX CC haemostatic, anti-HIV, antiasthmatic, antibacterial, virucide,
 XX CC neuroprotective, nontropic, antiparkinsonian, and antilipase activity.
 XX CC A polynucleotide encoding a polypeptide of the invention may have a use
 XX CC in gene therapy, and as a vaccine. A polypeptide of the invention is
 XX CC useful in the manufacture of a medicament for treating a syndrome
 XX CC associated with a human disease, the disease selected from a pathology
 XX CC associated with the polypeptide. These may also be used in diagnosing,
 XX CC treating or preventing NOVX-associated disorders such as cardiomyopathy,
 XX CC atherosclerosis, hypertension, scleroderma, obesity, cancer, diabetes,
 XX CC haemophilia, graft-versus-host disease, AIDS, asthma, Crohn's disease,
 XX CC multiple sclerosis, infections, anorexia, cancer-associated cachexia,
 XX CC neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's
 XX CC disease), haematopoietic disorders, dyslipidaemias and other wasting
 XX CC disorders associated with chronic diseases. The nucleic acids are also
 XX CC used as hybridisation probes, in chromosome mapping, tissue typing,
 XX CC preventive medicine, and pharmacogenomics. The polypeptides are also

CC useful as vaccines. The present sequence encodes a NOVX polypeptide of
 CC the invention.
 XX SQ Sequence 9698 BP; 1693 A; 3352 C; 3163 G; 1490 T; 0 U; 0 Other;
 XX
 XX Query Match 7.4%; Score 223.2; DB 10; Length 9698;
 XX Best Local Similarity 59.0%; Pred. No. 1.2e-30;
 XX Matches 466; Conservative 0; Mismatches 303; Indels 21; Gaps 4;
 XX
 XX QY 33 GGGAGCGAGACATCTCGCGCGGTGAGCCACCGCTGGTTCACGGGGTGTGGACCAAGT 92
 XX DB 4814 GGGAGGCGCGCTCTGCGCCAGGCTCCAGCAGCACTGTCTCTTCTTCCATGAGGCT 4873
 XX QY 93 TTGAGACCGGCAAGACCTCATCTCTCTGAGCTGTCTCATCCGAGGAGCTGCTGG 152
 XX DB 4874 TCGAGAGCGCGCGGAGCTGGTCAATGTTCACCGAGCTCTGCACA--GAGGAGCTGCTGG 4930
 XX QY 153 ACCGCTGTATCAGGAAGGGCGTGTGACGGAGGCGGAGGTCAAGTCTTACATCCAGCAGC 212
 XX DB 4931 AGCGAATCGCCAGGAACCCACCGTGTGTGAGTCTGAGATCCGGGCTATATCGGCAGG 4990
 XX QY 213 TGGTGGAGGGCTGCATCTACCTGCACAGCCATGCGGTTCTCCACTGGACATAAAGCCCT 272
 XX DB 4991 TGCTAGAGGGAATACACTACCTGCACAGAGCCACGCTGCACCTCGATGCAAGCCTG 5050
 XX QY 273 CTAACATCTGATGGTGCATCCTGC-----CCGGGAAGACATTAAATCTGCGACTTTG 326
 XX DB 5051 AGAACCTGCTGTGTGGATGTGCTCGGGCGAGCAGAGTGGGATCTGTGACTTTG 5110
 XX QY 327 GCTTTGCCAGAACATCACCCACAGAGCTGCAAGTTTCAGCCAGTACGGCTCCCTCGAGT 386
 XX DB 5111 GGAATGCCAGGAGCTGACTCCAGAGAGGCCAGTACTGCGAGTATGGCACCTCGAGT 5170
 XX QY 387 TCGTCTCCCCGAGATCATCCAGAGAACCTGTGTGAGGAGAACCTCCGACATTTGGGCCA 446
 XX DB 5171 TTGTAGCACCGAGATTGTCAATCAGAGCCCGTGTCTGGAGTCACTGACATCTGGCCTG 5230
 XX QY 447 TGGGTGTCACTCTCTACTGACCTGACCTGCTCATCCCCATTTCCCGCGGAGAGTGACC 506
 XX DB 5231 TGGGTGTGTGTGCTCTCTCTGTCTGACAGGAATCTCCCCGTTTGTGGGAAAATGACC 5290
 XX QY 507 GTGCCACCTCTCTGAAAGTCTCTGGAGGGCGCGTGTGTATGGAGCAGCCCCATGGCTGCC 566
 XX DB 5291 GGACAAATTGATGAACATCGAAACTACAAGTGGCTTCGAGAGAGCACATTCCTGA 5350
 XX QY 567 ACCTCAGCAGAGAGCCGCAAGACTTTCATCAAGGCTACGCTGCAGAGAGCCCTCAGGCC 626
 XX DB 5351 GCCTGAGCAGGAGGCGCGGGCTTCTCTCATCAAAAGTGTGGTGCAGGACCGGTG---A 5407
 XX QY 627 GGCCTAGTGGCGCCAGTGTCTCTCCACCCCTGTTCTCTGAAATCCATGCTGGGAGG 686
 XX DB 5408 GACCTACCGCAGAGAGACCTTAGAACATCTCTTGGTT-----CAAAACTCAGGCAG 5458
 XX QY 687 AGGCCCACTTCATCAACACCAAGCAGCTCAAGTCTCTCGCGCCGAGTGCCTGGCAGC 746
 XX DB 5459 AGGGCGCAGAGGTGAGCAGCGATCACTGAAAGTATTCCTCTCCGGCGGAGTGGCAGC 5518
 XX QY 747 GTTCCCTGATGAGCTACAAGTCCATCTCTGGTGTGTCGCTCCATCCCTGAGCTGTGCGG 806
 XX DB 5519 GCTCCAGATCAGCTACAAATGCCACCTGTGTGCTGGCGCCCATCCCCGAGCTGCTGCGG 5578
 XX QY 807 GCCCACCAGA 816
 XX DB 5579 CCCCCCAGAG 5588
 XX
 XX RESULT 15
 XX ADJ78945
 XX ID ADJ78945 standard; DNA; 9698 BP.
 XX XX
 XX AC ADJ78945;
 XX XX
 XX DT 06-MAY-2004 (first entry)

XX DE Human NOVX protein Nov14C gene sequence.
 XX KW NOVX; cytostatic; antidiabetic; anorectic; cerebroprotective;
 KW neuroprotective; antiinflammatory; thyromimetic; cardiant; gene-therapy;
 KW antisense-therapy; cancer; diabetes; obesity; endocrine disorder;
 KW CNS disorder; cardiovascular disorder; inflammatory disorder;
 KW detection assay; screening assay; chromosome mapping; tissue typing;
 KW predictive medicine; human; Nov14C; gene; ds.
 XX OS Homo sapiens.
 XX US US2004014053-A1.
 XX PD 22-JAN-2004.
 XX PF 01-AUG-2002; 2002US-00210130.
 XX PR 02-AUG-2001; 2001US-0309501P.
 PR 03-AUG-2001; 2001US-0310291P.
 PR 08-AUG-2001; 2001US-0310951P.
 PR 09-AUG-2001; 2001US-0311292P.
 PR 13-AUG-2001; 2001US-0311979P.
 PR 14-AUG-2001; 2001US-0312203P.
 PR 17-AUG-2001; 2001US-0313158P.
 PR 17-AUG-2001; 2001US-0313201P.
 PR 20-AUG-2001; 2001US-0313643P.
 PR 20-AUG-2001; 2001US-0313702P.
 PR 21-AUG-2001; 2001US-0314031P.
 PR 23-AUG-2001; 2001US-0314466P.
 PR 28-AUG-2001; 2001US-0315403P.
 PR 29-AUG-2001; 2001US-0315853P.
 PR 31-AUG-2001; 2001US-0316508P.
 PR 17-SEP-2001; 2001US-0322716P.
 PR 21-SEP-2001; 2001US-0323936P.
 PR 03-DEC-2001; 2001US-0338078P.
 PR 05-FEB-2002; 2002US-0354655P.
 PR 19-APR-2002; 2002US-0361764P.
 PR 15-MAY-2002; 2002US-0380971P.
 PR 15-MAY-2002; 2002US-0380980P.
 PR 16-MAY-2002; 2002US-0381039P.
 PR 28-MAY-2002; 2002US-0383761P.
 PR 29-MAY-2002; 2002US-0383887P.
 XX (ZERH/) ZERHUSEN B D.
 PA (PATT/) PATTURAJAN M.
 PA (KEKU/) KEKUDA R.
 PA (MILL/) MILLER C E.
 PA (RIEG/) RIEGER D K.
 PA (PENA/) PENNA C E A.
 PA (SHIM/) SHIMKETS R A.
 PA (LILL/) LI L.
 PA (BERG/) BERGHS C.
 PA (ZHON/) ZHONG M.
 PA (CASM/) CASMAN S J.
 PA (VOSS/) VOSS E Z.
 PA (BOLD/) BOLDOG F L.
 PA (FADI/) PADIGARU M.
 PA (SMIT/) SMITHSON G.
 PA (JIW/) JI W.
 PA (GORM/) GORMAN L.
 PA (VERN/) VERNET C A M.
 PA (LEIT/) LEITE M W.
 PA (GUOX/) GUO X S.
 PA (ANDE/) ANDERSON D W.
 PA (SPYT/) SPYTEK K A.
 PA (GERL/) GERLACH V.
 PA (BURG/) BURGESS C E.
 PA (KHRA/) KHRAMTSOV N V.
 PA (ORTT/) ORT T.
 PA (ELLE/) ELLERMAN K.
 PA (RAST/) RASTELLI L.

PA (AGEE/) AGEE M L.
 PA (CHAU/) CHAUDHURI A.
 PA (CHAN/) CHANT J S.
 PA (DIP/) DIPIPPO V A.
 PA (EDIN/) EDINGER S R.
 PA (EISE/) EISEN A J.
 PA (GANG/) GANGOLLI E A.
 PA (GIOT/) GIOT L.
 PA (OOIC/) OOI C E.
 PA (ROTH/) ROTHENBERG M E.
 PA (SPAD/) SPADERNA S K.
 PA (HJAL/) HJALT T.
 PA (LIUX/) LIU X.
 PA (TAUP/) TAUPIER R J.
 PA (CATT/) CATTERTON E.
 PA (SHEN/) SHENOY S G.
 XX ZERHUSEN BD, Patturajan M, Kekuda R, Miller CE, Rieger DK;
 PI Pena CE, Shinkets RA, Li L, Berghs C, Zhong M, Casman SJ, Voss EZ;
 PI Bollog FL, Padigaru M, Smithson G, Ji W, Gorman L, Vernet CM;
 PI Leite MW, Guo XS, Anderson DW, Spytek KA, Gerlach V, Burgess CE;
 PI Khrantsov NV, Ort T, Ellerman K, Rastelli L, Agee ML, Chaudhuri A;
 PI Chant JS, Dipippo VA, Edinger SR, Eisen AJ, Gangolli EA, Giot L;
 PI Ooi CE, Rothenberg ME, Spaderna SK, Hjalt T, Liu X, Taupier RJ;
 PI Catterton E, Shenoj SG;
 XX WPI; 2004-108206/11.
 DR P-PSDB; ADJ78946.
 XX PT New isolated NOVX polypeptides and nucleic acid molecules useful for
 PT treating, preventing and diagnosing pathological conditions with NOVX-
 PT associated disorders, such as cancer, obesity, diabetes and inflammatory
 PT or CNS diseases.
 XX Claim 20; SEQ ID NO 37; 250pp; English.
 CC This invention relates to a novel isolated NOVX polypeptide comprising a
 CC fully defined sequence of, a mature form, one or more conservative
 CC substitutions or at least 95% identity to 247 amino acids as given in the
 CC specification. The invention may be useful for the development of
 CC compounds with a cytostatic, antidiabetic, anorectic, cerebroprotective,
 CC neuroprotective, antiinflammatory, thyromimetic or cardiant activity. In
 CC addition, the disclosed sequences may prove useful for gene-therapy or
 CC antisense-therapy. The invention may be useful for the diagnosis and
 CC treatment of disorders associated with aberrant expression or activity of
 CC the NOVX polypeptide, such as cancer, diabetes, obesity, and endocrine,
 CC CNS, cardiovascular and inflammatory disorders. They can also be used in
 CC various detection and screening assays, chromosome mapping, tissue typing
 CC and predictive medicine. The present sequence is that of a gene which
 CC encodes a human NOVX protein of the invention.
 XX Sequence 9698 BP; 1693 A; 3352 C; 3163 G; 1490 T; 0 U; 0 Other;
 SQ
 Query Match 7.4%; Score 223.2; DB 12; Length 9698;
 Best Local Similarity 59.0%; Pred. No. 1.2e-30;
 Matches 466; Conservative 0; Mismatches 303; Indels 21; Gaps 4;
 QY 33 GGGAGCGAGACATCTGGCGGCTGAGCCACCCGCTGTCACGGGGTGTGGACCACT 92
 DB 4814 GGGAGCGCGGCTGTGGCCAGGCTCCAGACGACTGTGTCTTCTTCTTCCATGAGGCT 4873
 QY 93 TTGAGACCCGCAAGACCTCTCATCTCTGAGAGCTGTGCTCATCCGAGGAGTGTCTGG 152
 DB 4874 TCGAGAGCGCGGGGACTGTGCTCATTTGTCACCGAGCTCTGCACA---GAGGAGCTGTCTGG 4930
 QY 153 ACCGCTGTACAGGAGGGCGTGTGTCAGGAGGGCGGAGGTCAAGTCTACATCCAGCAGC 212
 DB 4931 AGCGAATCGCCAGGAAACCCACCGTGTGTGAGTCTGAGATCCGGGCGCTATATGGCGCAGG 4990
 QY 213 TGGTGGAGGGGCTGCACCTACCTGACAGCCATGCGGCTTCTCCACTGGACATAAGCCCT 272
 DB 4991 TGCTAGAGGGAATACACTACTCTCCACGAGCCAGCTGCTGCACCTCGATGTCAGGCTG 5050

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:15:10 ; Search time 639.09 Seconds
(without alignments)
10198.307 Million cell updates/sec

Title: US-10-077-130-6_COPY_13600_14700

Perfect score: 1101

Sequence: 1 cacagcagccactgtgac.....agcaggaaggcccatgttc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

- 1: Geneseqn1980a.*
- 2: Geneseqn1990a.*
- 3: Geneseqn2000a.*
- 4: Geneseqn2001a.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002a.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003a.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004a.*
- 13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1101	100.0	24120	8	ABX11642 Human ser
2	1076	97.7	20489	12	ADQ22881 Human ser
3	886.8	80.5	14061	6	ABV99363 Human NOV
4	886.8	80.5	14109	6	ABV99362 Human NOV
5	204.2	18.5	1494	10	ADF58317 Human pol
6	71.2	6.5	2000	8	ADA71938 Rice gene
7	53.4	4.9	2534	11	ADM03261 Human cDN
8	53.4	4.9	7564	8	ABZ24581 Human cel
9	50.2	4.6	1005	6	ABN21414 Human ORF
10	49.4	4.5	2000	8	ADA71938 Rice gene
11	48.6	4.4	1813	12	ADQ22659 Human sof
12	48.6	4.4	2167	13	ACN41796 Human dia
13	48.6	4.4	2167	13	ACN41509 Human dia
14	46.6	4.2	3613	12	ADQ21633 Human sof
15	46.6	4.2	4176	12	ADQ25448 Human sof
16	45.6	4.1	2010	1	AAN90362 Glucose i
17	45	4.1	2145	5	AAS66467 DNA encod
18	45	4.1	33529	5	AAS17367 DNA seque
19	45	4.1	67251	10	ADC28995 Sorangium
20	44.8	4.1	765	11	ABD16446 Pseudomon

c	21	44.8	4.1	1329	11	ABD16328	Abd16328 Pseudomon
c	22	44.8	4.1	5121	11	ABD16585	Abd16585 Pseudomon
c	23	44	4.0	735	11	ABD15635	Abd15635 Pseudomon
c	24	44	4.0	1410	11	ABD15421	Abd15421 Pseudomon
c	25	44	4.0	27705	10	ADC26979	Adc26979 Sorangium
c	26	43.8	4.0	815	6	AAS62129	Aas62129 Porcine m
c	27	43.8	4.0	817	6	AAS62126	Aas62126 Porcine m
c	28	43.8	4.0	819	6	AAS62125	Aas62125 Porcine m
c	29	43.8	4.0	820	6	AAS62127	Aas62127 Porcine m
c	30	43.8	4.0	825	6	AAS62128	Aas62128 Porcine m
c	31	43.8	4.0	1038	13	ADS55903	AdS55903 Bacterial
c	32	43.8	4.0	9590	8	ACA45481	AcA45481 Prokaryot
c	33	43.8	4.0	177587	11	ACN44806	Acn44806 Human gen
c	34	43.6	4.0	2737	4	AAI59251	Aai59251 Human pol
c	35	43.6	4.0	3999	4	AAI61037	Aai61037 Human pol
c	36	43.6	4.0	5378	13	ADR67197	Adr67197 Human bla
c	37	43.6	4.0	5382	10	ADD14722	Add14722 Human bla
c	38	43.4	3.9	1338	12	ADP67802	Adp67802 Rice G339
c	39	43.2	3.9	1578	12	ACH87202	Ach87202 Human gen
c	40	43.2	3.9	3015	13	ADQ86796	Adq86796 Human tum
c	41	43.2	3.9	3122	4	AAH14898	Aah14898 Human cDN
c	42	43.2	3.9	3340	5	AAS45087	Aas45087 cDNA enco
c	43	43	3.9	3131	4	ABL04502	Abi04502 Drosophil
c	44	43	3.9	3131	4	ABL20422	Abi20422 Drosophil
c	45	43	3.9	5304	4	ABL20423	Abi20423 Drosophil

ALIGNMENTS

RESULT 1

ABX11642

ID ABX11642 standard; cDNA; 24120 BP.

AC ABX11642;

DT 09-MAY-2003 (first entry)

DE Human serine/threonine or protein kinase 12599, cDNA.

KW Human; ss; Gene; serine/threonine kinase; protein kinase; 12599;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.

XX Homo sapiens.

OS Homo sapiens.

Key Location/Qualifiers

5'UTR 1..71

CDS /tag= a

72..23978 /tag= b

/product= "Kinase 12599"

/note= "This CDS is specifically claimed in claim 2"

3'UTR 23979..24120

/tag= c

US2002168742-A1.

14-NOV-2002.

15-FEB-2002; 2002US-00077130.

15-FEB-2001; 2001US-0269201P.

(MILL-) MILLENNIUM PHARM INC.

Kapeller-Libermann R, Acton SL;

XX

DR WPI; 2003-298729/29.
XX P-PSDB; ABG76187.
PT Novel isolated human protein kinase, designated 59079 or 12599
PT polypeptide, useful as diagnostic and therapeutic agents for preventing
PT cardiovascular diseases, proliferative disorders, and protein kinase
PT disorders.
XX
PS Claim 2; Page 58-84; 119pp; English.
XX
CC The invention relates to an isolated human serine/threonine or protein
CC kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
CC comprising at least 85% identity to the nucleic acids appearing as
CC ABX11641 and ABX11642 or their complement, a naturally occurring variant
CC of the kinases or their fragments. Also included are a non-human host
CC cell containing the nucleic acids, an antibody specific for the proteins,
CC identifying a compound which binds to the kinase (by contacting the
CC kinase or a cell expressing the kinase with a test compound and
CC determining whether the kinase binds to the test compound) and modulating
CC the activity of kinase using the identified compound. The kinases and
CC their encoding nucleic acids are useful as diagnostic and therapeutic
CC agents for preventing a disease or condition associated with an aberrant
CC or unwanted 59079 or 12599 activity in a subject, including
CC cardiovascular diseases such as heart failure, and myocardial infarction;
CC disorders involving blood vessels such as atherosclerosis, and Kaposi's
CC sarcoma; blood platelets disorder such as thrombocytopaenia, and Kaposi's
CC Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders
CC such as cancer; and protein kinase disorders such as autoimmune
CC disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,
CC rheumatoid arthritis, and multiple sclerosis (many examples of diseases
CC and disorders are included in the specification). The kinases, their
CC encoding nucleic acids and antibodies are useful in screening assays,
CC detection assays (e.g. forensic biology), and predictive medicine (e.g.
CC diagnostic assays, prognostic assays, and monitoring clinical trials and
CC pharmacogenomics). The kinases and their encoding nucleic acids are
CC useful as query sequences to perform a search against public databases to
CC identify other family members or related sequences. The present sequence
CC encodes the kinase 12599
XX
SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;

Query Match 100.0%; Score 1101; DB 8; Length 24120;
Best Local Similarity 100.0%; Pred. No. 5.5e-233;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACAGACCCACACTGTGACACTGTCTTGGCAGCTCCCATGATGATGGAGCGGTGGT 13730
DB 13671 CACAGACCCACACTGTGACACTGTCTTGGCAGCTCCCATGATGATGGAGCGGTGGT 13730

QY 61 CTCTGTGGCTACCGGTGAGGTGAGAGAGGGGGCCACAGGCCAGTGGCGCTGTGCCAC 120
DB 13731 CTCTGTGGCTACCGGTGAGGTGAGAGAGGGGGCCACAGGCCAGTGGCGCTGTGCCAC 13790

QY 121 GAGCTGTGTGCTGACCCGAGTGTGTGGTATGCTTGGCTGGCCCGGGGAGACCTACCGC 180
DB 13791 GAGCTGTGTGCTGACCCGAGTGTGTGGTATGCTTGGCTGGCCCGGGGAGACCTACCGC 13850

QY 181 TTCCTGTGGCAGTGTGGCCCTGTGGTGTCTGGGAAACCGGTTACCTGCCCCAGACA 240
DB 13851 TTCCTGTGGCAGTGTGGCCCTGTGGTGTCTGGGAAACCGGTTACCTGCCCCAGACA 13910

QY 241 GTGGCGCTTGCAGAGCCACAGCCTGTGCTTCCCGAGCCCTCAGCCCTTGAGAGCCG 300
DB 13911 GTGGCGCTTGCAGAGCCACCGAGCCTGTGCTTCCCGAGCCCTCAGCCCTTGAGAGCCG 13970

QY 301 CAGTGTGCAGTGTGAGATGTCTCTGAGGCTTGAGGTGGGTGAGGTGGTGGAG 360
DB 13971 CAGTGTGCAGTGTGAGATGTCTCTGAGGCTTGAGGTGGGTGAGGTGGTGGAG 14030

QY 361 GTCATCTGGCACAAGGAATGAGCGCATCCAGCCCGTGGCGGTTCGAGGTGGTCTCC 420
DB 14031 GTCATCTGGCACAAGGAATGAGCGCATCCAGCCCGTGGCGGTTCGAGGTGGTCTCC 14090

QY 421 CAGGTCGGCAACAGATGCTGTGATCAAGGGCTTTCACGGCAGAAAGACCCAGGCGAGTAC 480
DB 14091 CAGGTCGGCAACAGATGCTGTGATCAAGGGCTTTCACGGCAGAAAGACCCAGGCGAGTAC 14150

QY 481 CACTGTGGCTGTGGCTCAGGGCTCCATCTGCGCTCGGGTGCACCTTTCAGGTGGCACTG 540
DB 14151 CACTGTGGCTGTGGCTCAGGGCTCCATCTGCGCTCGGGTGCACCTTTCAGGTGGCACTG 14210

QY 541 AGCCCAAGCTCTGTGGATGAGGCCCTTCAGCCCAAGCTTGGCCCGGAGGAGCCAGGAG 600
DB 14211 AGCCCAAGCTCTGTGGATGAGGCCCTTCAGCCCAAGCTTGGCCCGGAGGAGCCAGGAG 14270

QY 601 GGTGACCTGCACCTACTGTGGAGGCCCTGTGCTCGGAAACGTTCGCATGAGCCGAGCC 660
DB 14271 GGTGACCTGCACCTACTGTGGAGGCCCTGTGCTCGGAAACGTTCGCATGAGCCGAGCC 14330

QY 661 ACGTGTGACTCCATTAGCGAGCTGCCAGAGAGGACCGGCCCTCGCAGCGCTTGCACAG 720
DB 14331 ACGTGTGACTCCATTAGCGAGCTGCCAGAGAGGACCGGCCCTCGCAGCGCTTGCACAG 14390

QY 721 GAGGCAGAGAGGTGGCACCTGATCTCTCTGAGGGCTACTCCACGGCGGATGAGCTGGCC 780
DB 14391 GAGGCAGAGAGGTGGCACCTGATCTCTCTGAGGGCTACTCCACGGCGGATGAGCTGGCC 14450

QY 781 CGCAGCTGGAGATGCTGACCTCTTCACACACAGCTCTCTGATGATGATGATGATGATGAT 840
DB 14451 CGCAGCTGGAGATGCTGACCTCTTCACACACAGCTCTCTGATGATGATGATGATGATGAT 14510

QY 841 CTTTCCCTGTGCTACCTTACCTCAAGAGGCTGGGAGGCGCAGGACCTTCCACCTGGCCAGC 900
DB 14511 CTTTCCCTGTGCTACCTTACCTCAAGAGGCTGGGAGGCGCAGGACCTTCCACCTGGCCAGC 14570

QY 901 AAGGTTGGGGCCCCCAGCAGCCCCCTCTGTGAAGCCACAGCAGCAGCAGCAGCAGCAGCAG 960
DB 14571 AAGGTTGGGGCCCCCAGCAGCCCCCTCTGTGAAGCCACAGCAGCAGCAGCAGCAGCAGCAG 14630

QY 961 GCTGTGGCCGCCACCTACCTGGGAGACCTGAGCACCACAAAGACCTGGGTGATCCCTCAATGGAC 1020
DB 14631 GCTGTGGCCGCCACCTACCTGGGAGACCTGAGCACCACAAAGACCTGGGTGATCCCTCAATGGAC 14690

QY 1021 AAGGCAGCTGTGAAGATCCAGGCTGGCTTTAAGGGCTTACAGGTCCTCGGAAGGAGATGAAG 1080
DB 14691 AAGGCAGCTGTGAAGATCCAGGCTGGCTTTAAGGGCTTACAGGTCCTCGGAAGGAGATGAAG 14750

QY 1081 CAGCAGAAAGGGGCCCATGTC 1101
DB 14751 CAGCAGAAAGGGGCCCATGTC 14771

RESULT 2

ADQ22881

ID ADQ22881 standard; DNA; 20489 BP.

XX AC ADQ22881;

XX AC ADQ22881;

DT 26-AUG-2004 (first entry)

XX

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5701.

XX

KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW db.

OS Homo sapiens.

XX WO2004048938-A2.

PN

PD 10-JUN-2004.

XX

PF 26-NOV-2003; 2003WO-US038193.

XX

PR 26-NOV-2002; 2002US-0429739P.

XX

PA (PROT-) PROTEIN DESIGN LABS INC.

XX PI Aziz N, Gineburg WM, Zlotnik A;
XX DR WPI; 2004-441208/41.
XX PT Early detection of soft tissue sarcoma comprises determining expression
XX PT of a gene in a first soft tissue sample and a normal soft tissue sample
XX PT and comparing the gene expression, also useful in treating soft tissue
XX PT sarcoma.
XX PS Example 2; SEQ ID NO 5701; 210pp; English.
XX CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX SQ Sequence 20489 BP; 3917 A; 5815 C; 6983 G; 3225 T; 0 U; 549 Other;
Query Match 97.7%; Score 1076; DB 12; Length 20489;
Best Local Similarity 97.7%; Pred. No. 1.8e-227;
Matches 1076; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 1 CACAGCAGCCACTGTGACACTGTCTTGGGAGCTTCCCATGATGATGGAGCGGTGGT 60
DB 13677 CACAGCAGCCACTGTGACACTGTCTTGGGAGCTTCCCATGATGATGGAGCGGTGGT 13736
QY 61 CTCTGTGCTACCGGTGGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGTGGCCAC 120
DB 13737 CTCTGTGCTACCGGTGGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGTGGCCAC 13796
QY 121 GAGCTGGTGCCTGGACCCGAGTGTGTGGTGGCTGGCCCTGGCCCGGGAGACCTACCGC 180
DB 13797 GAGCTGGTGCCTGGACCCGAGTGTGTGGTGGCTGGCCCTGGCCCGGGAGACCTACCGC 13856
QY 181 TTCCGTGTGGCAGCTGTGGGCGCTGTGGTGTGGGGAACCGGTTCACTGCCCCAGACA 240
DB 13857 TTCCGTGTGGCAGCTGTGGGCGCTGTGGTGTGGGGAACCGGTTCACTGCCCCAGACA 13916
QY 241 GTGGCGCTTGACAGCCACCGAGCTGTGCTTCCCGACCTCAGCCCCCTGAGAGCCGG 300
DB 13917 GNN 13976
QY 301 CAGGTGGCAGCTGGTGAAGATGTCTCTGGAGCTTGAGGTGGTGGCTGAGGCTGGTGAG 360
DB 13977 CAGGTGGCAGCTGGTGAAGATGTCTCTGGAGCTTGAGGTGGTGGCTGAGGCTGGTGAG 14036
QY 361 GTCACTCTGGCAAGGGAATGAGCGCATCCAGCCCGGTGGGCGGTTCCAGGTGGTCTCC 420
DB 14037 GTCACTCTGGCAAGGGAATGAGCGCATCCAGCCCGGTGGGCGGTTCCAGGTGGTCTCC 14096
QY 421 CAGGTGGCAGCATGCTGGTGAATCAAGGCTTCAGGGCAGAGACACGGGCGAGTAC 480
DB 14097 CAGGTGGCAGCATGCTGGTGAATCAAGGCTTCAGGGCAGAGACACGGGCGAGTAC 14156
QY 481 CACTGTGGCTGGCTCAGGGCTCCATCTGCGCTGGGCTGGCCACCTTCAGGTGGCACTG 540
DB 14157 CACTGTGGCTGGCTCAGGGCTCCATCTGCGCTGGGCTGGCCACCTTCAGGTGGCACTG 14216
QY 541 AGCCGAGCCTCTGTGGATGAGGCCCTCAGCCCCAGCTTTCGCCCGGAGGCCAGCCAGGAG 600
DB 14217 AGCCGAGCCTCTGTGGATGAGGCCCTCAGCCCCAGCTTTCGCCCGGAGGCCAGCCAGGAG 14276
QY 601 GGTGACCTGCACCTACTACTGTGGAGGCCCTTGGCTCGGAAACGTCGCGATGAGCGGTAGGCC 660

DB 14277 GGTGACCTGCACCTACTGTGGAGGCCCTTGGCTCGGAAACGTCGCGATGAGCGGTAGGCC 14336
QY 661 ACCTGTGACTCCATTTAGCGAGCTGCCAGAGGAGGACGGCCGCTCGCAGCGCTGCCACAG 720
DB 14337 ACCTGTGACTCCATTTAGCGAGCTGCCAGAGGAGGACGGCCGCTCGCAGCGCTGCCACAG 14396
QY 721 GAGCAGAGGAGGTGGCAGCTGTCTCTGAAGGCTACTCCAGCGGCGATGAGTGGCC 780
DB 14397 GAGCAGAGGAGGTGGCAGCTGTCTCTGAAGGCTACTCCAGCGGCGATGAGTGGCC 14456
QY 781 CGCACTGGAGATGCTGACCTCTCACACACAGCTCTGATGATGAGTCCCGGGCAGGCACC 840
DB 14457 CGCACTGGAGATGCTGACCTCTCACACACAGCTCTGATGATGAGTCCCGGGCAGGCACC 14516
QY 841 CTTTCCCTGGTCACTACCTCAAGAAGGCTGGGAGGCCAGGCACCTCACCACTGGCCAGC 900
DB 14517 CTTTCCCTGGTCACTACCTCAAGAAGGCTGGGAGGCCAGGCACCTCACCACTGGCCAGC 14576
QY 901 AAGTTGGGGCCCCAGCAGCCCTCTGTGAAGCCACAGCAGCAGCAGGACCCACTGGCT 960
DB 14577 AAGTTGGGGCCCCAGCAGCCCTCTGTGAAGCCACAGCAGCAGCAGGAGCCACTGGCT 14636
QY 961 GCTGTGCCCCACCACTGGGAGACTGAGCAGCACCAGACCTGGGTGATCCCTCAATGAC 1020
DB 14637 GCTGTGCCCCACCACTGGGAGACTGAGCAGCACCAGACCTGGGTGATCCCTCAATGAC 14696
QY 1021 AAGCAGCTGTGAAGATCCAGGCTGCCCTTTAAGGGCTACAAGGTCGGAAGAGATGAAG 1080
DB 14697 AAGCAGCTGTGAAGATCCAGGCTGCCCTTTAAGGGCTACAAGGTCGGAAGAGATGAAG 14756
QY 1081 CAGCAGGAGGGCCCCATGTTT 1101
DB 14757 CAGCAGGAGGGCCCCATGTTT 14777
RESULT 3
ID ABV99363 standard; DNA; 14061 BP.
XX AC ABV99363;
XX DT 27-JAN-2003 (first entry)
XX DE Human NOV13b coding sequence.
XX KW Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS;
KW antiinflammatory; cardiac; haemostatic; neuroprotective; anorectic;
KW neutropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;
KW antifertility; cerebroprotective; gene therapy; NOVA; NOV; fertility;
KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;
KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; cardiovascular disorder;
KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;
KW metabolic syndrome X; wasting disorder; cell differentiation; gene;
XX OS Homo sapiens.
XX PN WO200272771-A2.
XX PD 19-SEP-2002.
XX PF 08-MAR-2002; 2002WO-US007288.
XX PR 08-MAR-2001; 2001US-0274101P.
PR 08-MAR-2001; 2001US-0274194P.
PR 08-MAR-2001; 2001US-0274281P.
PR 09-MAR-2001; 2001US-0274832P.
PR 12-MAR-2001; 2001US-0275235P.
PR 13-MAR-2001; 2001US-0275578P.
PR 13-MAR-2001; 2001US-0275579P.

PR 13-MAR-2001; 2001US-0275601P.
 PR 14-MAR-2001; 2001US-0276000P.
 PR 16-MAR-2001; 2001US-0276776P.
 PR 19-MAR-2001; 2001US-0276994P.
 PR 20-MAR-2001; 2001US-0277239P.
 PR 20-MAR-2001; 2001US-0277321P.
 PR 20-MAR-2001; 2001US-0277327P.
 PR 20-MAR-2001; 2001US-0277338P.
 PR 21-MAR-2001; 2001US-0277791P.
 PR 22-MAR-2001; 2001US-0277833P.
 PR 23-MAR-2001; 2001US-0278152P.
 PR 26-MAR-2001; 2001US-0278894P.
 PR 27-MAR-2001; 2001US-0278999P.
 PR 27-MAR-2001; 2001US-0279036P.
 PR 28-MAR-2001; 2001US-0279344P.
 PR 30-MAR-2001; 2001US-0279959P.
 PR 30-MAR-2001; 2001US-0280233P.
 PR 02-APR-2001; 2001US-0280802P.
 PR 02-APR-2001; 2001US-0280822P.
 PR 02-APR-2001; 2001US-0280900P.
 PR 04-APR-2001; 2001US-0281194P.
 PR 13-APR-2001; 2001US-0283675P.
 PR 30-APR-2001; 2001US-0287424P.
 PR 02-MAY-2001; 2001US-0288066P.
 PR 03-MAY-2001; 2001US-0288344P.
 PR 03-MAY-2001; 2001US-0288528P.
 PR 15-MAY-2001; 2001US-0291190P.
 PR 16-MAY-2001; 2001US-0291099P.
 PR 16-MAY-2001; 2001US-0291240P.
 PR 30-MAY-2001; 2001US-0294485P.
 PR 31-MAY-2001; 2001US-0294889P.
 PR 31-MAY-2001; 2001US-0294899P.
 PR 18-JUN-2001; 2001US-0299027P.
 PR 19-JUN-2001; 2001US-0299303P.
 PR 19-JUN-2001; 2001US-0299310P.
 PR 10-JUL-2001; 2001US-0304354P.
 PR 31-JUL-2001; 2001US-0309198P.
 PR 16-AUG-2001; 2001US-0312903P.
 PR 10-SEP-2001; 2001US-0318462P.
 PR 12-SEP-2001; 2001US-0318770P.
 PR 27-SEP-2001; 2001US-0325430P.
 PR 18-OCT-2001; 2001US-0325681P.
 PR 18-OCT-2001; 2001US-0330380P.
 PR 31-OCT-2001; 2001US-0335301P.
 PR 14-NOV-2001; 2001US-0332172P.
 PR 14-NOV-2001; 2001US-0332271P.
 PR 14-NOV-2001; 2001US-0332272P.
 PR 14-NOV-2001; 2001US-0333184P.
 PR 14-NOV-2001; 2001US-0333272P.
 PR 21-NOV-2001; 2001US-0332094P.
 PR 03-DEC-2001; 2001US-0337426P.
 PR 03-DEC-2001; 2001US-0338092P.
 PR 04-DEC-2001; 2001US-0337185P.
 PR 03-JAN-2002; 2002US-0345705P.
 PR 08-MAR-2002; 2002US-00093463.
 PR (CURA-) CURAGEN CORP.
 PR Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
 PI Boldog FL, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Vernet CAM;
 PI Pena CE, Burges CE, Liu X, Spytek KA, Gorman L, Spaderna SK;
 PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
 PI Taupier RJ, Padigar M, Shenoy SG, Kekuda R, Gusev VY, Pochart PF;
 PI Zhong M;
 PR WPI; 2002-732824/79.
 DR P-PSDB; ABP70085.
 PR XX
 PR The present invention relates to new isolated proteins (NOVX) and their
 CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
 CC any number from 1 to 48. The NOVX proteins and coding sequences are
 CC useful in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, preferably a NOVX-associated disorder.
 CC The NOVX coding sequences and proteins are useful for treating,
 CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
 CC obesity, infectious disease, anorexia, cancer-associated cachexia,
 CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
 CC disease, immune disorders, haematopoietic disorders, cardiovascular
 CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
 CC disturbances associated with obesity, metabolic syndrome X or wasting
 CC disorders associated with chronic diseases or various cancers. The NOVX
 CC coding sequences and proteins may also be used as targets for the
 CC identification of small molecules that modulate or inhibit e.g.
 CC neurogenesis, cell differentiation, cell proliferation, haematopoiesis,
 CC wound healing and angiogenesis, in gene therapy, in generation of
 CC antibodies that bind immunospecifically to NOVX substances for use in
 CC therapeutic or diagnostic methods
 PR XX
 PR Sequence 14061 BP; 2854 A; 4029 C; 4839 G; 2339 T; 0 U; 0 Other;
 Query Match 80.5%; Score 886.8; DB 6; Length 14061;
 Best Local Similarity 98.9%; Pred. No. 9.2e-186;
 Matches 904; Conservative 0; Mismatches 7; Indels 3; Gaps 1;
 QY 1 CACAGCAGCCACACTGTGACACTGTCTTGGGCGAGCTCCCATGAGTGTGAGGCGGTGGT 60
 DB 9366 CGCAGCAGCCACACTGTGACACTGTCTTGGGCGAGCTCCCATGAGTGTGAGGCGGTGGT 9425
 QY 61 CTCTGTGGCTACCGCGTGGAGGTGAAGAGGCGGCGCACAGCCAGTGGCGGTGTGCCAC 120
 DB 9426 CTCTGTGGCTACCGCGTGGAGGTGAAGAGGCGGCGCACAGCCAGTGGCGGTGTGCCAC 9485
 QY 121 GAGCTGTGCTGGACCCGAGTGTGTGGTGGTGGCTGGCCCTCCCGGGGAGACTTACCGC 180
 DB 9486 GAGCTGTGCTGGACCCGAGTGTGTGGTGGTGGCTGGCCCTCCCGGGGAGACTTACCGC 9545
 QY 181 TTCCGTGTGGCAGCTGTGGGCGCTGTGGGTGTGGGAGACCGGTTCACCTGCCCGCAGACA 240
 DB 9546 TTCCGTGTGGCAGCTGTGGGCGCTGTGGGTGTGGGAGACCGGTTCACCTGCCCGCAGACA 9605
 QY 241 GTGCGGCTTGCAGAGCCACCGAAGCTGTGCTCCCTCCCGAGCCCTCAGCCCTTGAGAGCCGG 300
 DB 9606 GTGCGGCTT---GAGCCACCGAAGCTGTGCTCCCTCCCGAGCCCTCAGCCCTTGAGAGCCGG 9662
 QY 301 CAGGTGGCAGCTGGTGAAGATGTCTCTGGAGCTTGAAGTGTGGTGTGGTGTGGTGTGGTGTGG 360
 DB 9663 CAGGTGGCAGCTGGTGAAGATGTCTCTGGAGCTTGAAGTGTGGTGTGGTGTGGTGTGGTGTGG 9722
 QY 361 GTCATCTGGCACAAGGGGATGGAGCCATCCAGCCCGGTGGGCGGTTCGAGGTGGTCTCC 420
 DB 9723 GTCATCTGGCACAAGGGGATGGAGCCATCCAGCCCGGTGGGCGGTTCGAGGTGGTCTCC 9782
 QY 421 CAGGTCGCGCAACAGATGTGTGTGATCAAGGGGTTTCAAGGCAAGAACACAGGCGCGAGTAC 480
 DB 9783 CAGGTCGCGCAACAGATGTGTGTGATCAAGGGGTTTCAAGGCAAGAACACAGGCGCGAGTAC 9842
 QY 481 CACTGTGGCTGTGGCTCAGGGCTCCATCTGCTGCTGGGCTGGCCACCTTCAGAGTGGCACTG 540
 DB 9843 CACTGTGGCTGTGGCTCAGGGCTCCATCTGCTGCTGGGCTGGCCACCTTCAGAGTGGCACTG 9902
 QY 541 AGCCAGGCTCTGTGGATGAGGCGCTCAGCCCTCAGCCAGCTTCCCGCCGAGCAGCCAGGAG 600
 DB 9903 AGCCAGGCTCTGTGGATGAGGCGCTCAGCCCTCAGCCAGCTTCCCGCCGAGCAGCCAGGAG 9962
 QY 601 GGTGACCTGCACCTACTGTGGGAGGCGCTGGGTTCGGAACGTCGCAATGAGCGGTGAGCCC 660
 DB 9963 GGTGACCTGCACCTACTGTGGGAGGCGCTGGGTTCGGAACGTCGCAATGAGCGGTGAGCCC 10022
 QY 661 ACCTGTGACTCCATTAGCGAGCTGCCAGAGGAGGAGCGGCGCTCGCAGCGCTGCCACAG 720

XX The invention relates to a novel human polynucleotide and the encoded
 CC polypeptide. A polynucleotide of the invention AD06202-AD06773 is useful
 CC therapy. An oligonucleotide of the invention AD06202-AD06773 is useful
 CC as a primer for synthesizing the polynucleotide or as a probe for
 CC detecting the polynucleotide. The polynucleotides AD06136-AD06378 are
 CC useful in gene therapy, for developing a diagnostic marker or medicines
 CC for regulating their expression and activity, or as a target of gene
 CC therapy. The proteins AD03759-AD06201 encoded by the polynucleotides
 CC are useful as pharmaceutical agents. The present sequence represents a
 CC cDNA sequence of the invention.

XX
 XX
 XX Sequence 2534 BP; 551 A; 679 C; 778 G; 526 T; 0 U; 0 Other;
 SQ

Query Match 4.9%; Score 53.4; DB 11; Length 2534;
 Best Local Similarity 53.6%; Pred. No. 0.053;
 Matches 111; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 287 CCCCTGAGAGCGGCGAGGTGGCAGCTGTGTGAAGATGTTCTCTGAGCTTGAGTGTGG 346
 DB 803 CCTGTGAAGACCACTGGTGGCGCCAGCGGAGGAGCTGGAGCTGGCTGTGAGCTGTAC 862

QY 347 CTGAGCTGGTGAAGTCTCTGCGACAGGGAATGAGCGCATCCAGCCCGTGGCGGT 406
 DB 863 GGGCGGGAACGCCCTGTGACTGCTGTGAAGGACAGGAGGCCATCCGCAAGGCCAGAA 922

QY 407 TCGAGGTGTTCTCCAGGGTCCGCAACAGATGCTGTGATCAAGGGCTTCACGGCAGA 466
 DB 923 ATGATGTGTTCTGCGAGGCGACGATGGCCATGCTGTGATCCGGGGGCTTCGCTCA 982

QY 467 ACCAGGCGAGTACCACTGTGGCCCTGG 493
 DB 983 ACGCGGGGAGTACACGCTGTGAGTGG 1009

RESULT 9
 ABZ24581
 ID ABZ24581 standard; cDNA; 7564 BP.
 AC ABZ24581;
 XX
 XX 31-MAR-2003 (first entry)
 DT
 XX Human cell adhesion and extracellular matrix protein 4 cDNA.
 KW Cell adhesion and extracellular matrix protein 4; CADECM-4; human;
 KW anti-HIV; virucide; anti-allergic; anti-inflammatory; antianemic;
 KW anti-parkinsonian; nootropic; anticonvulsant; anti-infective;
 KW anti-arteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
 KW cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;
 KW angiot; thyromimetic; neuroprotective; osteoprotective; antiarthritic;
 KW antiparasitic; antihelminthic; antiparasitic; uropathic; ophthalmological;
 KW antirheumatic; haemostatic; antibacterial; protozoacide; fungicide;
 KW gynaecological; titin; gene therapy; gene; ss.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH 241..7227
 CDS /*tag= a
 FT /product= "Human CADECM-4"
 FT
 FT
 FT
 FT
 FN WO200288322-A2.
 XX
 XX 07-NOV-2002.
 XX
 XX 01-MAY-2002; 2002WO-US013874.
 XX
 XX 02-MAY-2001; 2001US-0288290P.
 PR 21-MAY-2001; 2001US-0292468P.
 PR 15-JUN-2001; 2001US-0298616P.
 PR 28-JUN-2001; 2001US-0301672P.
 PR 04-JAN-2002; 2002US-0345008P.

XX (INCY-) INCYTE GENOMICS INC.
 XX Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding L;
 PI Hillman JL, Baughn MR, Kallick DA, Lee S, Warren BA, Xu Y, Tran UK;
 PI Lal PG, Thornton M, Hafalia AJA, Yao MG, Nguyen DB, Gandhi AR;
 PI Khan FA, Walia NK, Griffin JA, Chinn AM, Elliott VS, Ramkumar J;
 XX Arvizu CS, Forsythe IU;
 DR WPI; 2003-167112/16.
 DR P-PSDB; ABP58227.
 XX
 XX New human cell adhesion and extracellular matrix proteins, useful for
 PT diagnosing, treating or preventing autoimmune or inflammatory disorder
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
 PT cancer or hepatitis.
 PS Claim 12; Page 168-170; 178pp; English.
 XX
 XX The present sequence is that of Incyte clone 7326129CB1 cDNA encoding
 CC human cell adhesion and extracellular matrix protein 4 (CADECM-4). A
 CC representative cDNA library for the full-length polynucleotide is
 CC MUSITR02, constructed from the muscle tissue RNA of a Caucasian adult
 CC man. Homology data suggest the encoded protein to a titin muscle protein.
 CC The invention provides CADECM-1 to -11 polypeptides (see ABP58224-34) and
 CC polynucleotides (see ABZ24578-88), expression vectors, host cells,
 CC antibodies, agonists and antagonists. These are useful for diagnosing,
 CC treating or preventing disorders associated with aberrant expression of
 CC CADECM, particularly cell proliferative disorders (e.g. arteriosclerosis,
 CC atherosclerosis, cirrhosis, hepatitis, paroxysmal nocturnal
 CC haemoglobinuria, polycythaemia vera, psoriasis, primary
 CC tubular acidosis, anaemia or mental retardation), neurological disorders
 CC (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), reproductive
 CC disorders (e.g. infertility or a disruption in the menstrual cycle), or
 CC autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma, autoimmune
 CC thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus,
 CC glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease,
 CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
 CC osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid
 CC arthritis, Sjogren's syndrome, uveitis), or viral, bacterial, fungal,
 CC parasitic, protozoal or helminthic infections
 XX
 SQ Sequence 7564 BP; 1594 A; 1999 C; 2711 G; 1260 T; 0 U; 0 Other;
 Query Match 4.9%; Score 53.4; DB 8; Length 7564;
 Best Local Similarity 53.6%; Pred. No. 0.065;
 Matches 111; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 287 CCCCTGAGAGCGGCGAGGTGGCAGCTGTGTGAAGATGTTCTCTGAGCTTGAGTGTGG 346
 DB 2612 CCCTGGAAGACCACTGGTGGCGCCAGGAGGAGACGTGGAGCTGCGCTGTGACTGT 2671

QY 347 CTGAGGTGTTGAGTGTGATCTGGCAAGGGAATGAGGCGCATCCAGCCCGTGGCGGT 406
 DB 2672 GGGCGGGAACGCCCTGTGCTGCTGCAAGGAGGAGGCAATCCGCAAGAGCCAGA 2731

QY 407 TCGAGGTGTTCTCCAGGGTTCGCAACAGATGCTGTGATCAAGGGCTTCACGGCAGA 466
 DB 2732 ATGATGTGTTCTGCGAGGGCACGATGGCCATGCTGTGATCCTCGCGGGGCTCGCTCA 2791

QY 467 ACCAGGCGGAGTACCACTGTGGCGCTGG 493
 DB 2792 ACGCGGCGAGTACACGCTGTGAGTGG 2818

RESULT 9
 ABN21414
 ID ABN21414 standard; cDNA; 1005 BP.
 XX
 XX ABN21414;
 XX
 XX 24-JUN-2002 (first entry)
 DT

QY	383	AGCGCATCCAGCCCGGTGGCGGTTTCGAGGTGGTCTCCAGGGTCCGCAACAGATGCTG	441
Db	585	SKMGRWSGMSRMYRMRWKKMRKRKYMRWKKWCTWRRCMCYRWGYTMYTTSRSRMYTGR	526
QY	443	TGATCAAGGGCTTCACGGCAGAGACACAGGGCGAGTACCACCTGTGGCTGGCTCAGGGCT	502
Db	525	YKARYTSKRRTYMYKYRKYCWYYYGYMYWKCYSMMRYGYCKACKCKCYAMCWAAYSGMW	466
QY	503	CCATCTGCCTCGGGCTGCACCTTCACAGTGGCACTGAGCCACAGCTCTGTGGATGAGG	562
Db	465	MYWRYKYSKWMRMSITKYMSMWYKCRSMKYGAKGCGYCKRMWYCSYGMYKWTYMGSYK	406
QY	563	C--CCCTAGCCAGCTTGCCCCCGGAGCAGCCAGGAGGTGACCTGCACCTACTGTG	620
Db	405	YGRCYKYMRYMYKGMWYMYYSAYSSMMTWYTYAKYKYYWYKRRGTMSWYKSYKKY	346
QY	621	GGAGGCCCTGGCTCGGAAACGTGCGATGAGCGCTGAGCCACGCTGGACTCCATTAGCGA	680
Db	345	CTWWCYMKCMRYRWKMKRKTYSKRCYCWRYATCYWCCCYRKGWYSRRSMRMTAG	286
QY	681	GCTGCAGAGGAGGACGGCCGCTCGCAGCGCCTGCACAGGAGGAGGAGGTGGCACC	740
Db	285	WKMRWSRWCRSYSWYKWKWKYSYMSYGWRSSTWRSAAKPTYKGYSTSRRAWM	226
QY	741	TGATCTCTGAAGCTACTCACGGCCGATAGCTGGCCCGCCTCGAGAGTGTGACCT	800
Db	225	WKAARMYSACRRYSRTSYCGCSYCGSKWYMSKSCSMRMTCSWCSCCYTCYYGAMCW	166
QY	801	CTCACACACCACTCTGATGATGAGTCCCGGSCAGGCACCCCTTCCCTGGTCACCTACCT	860
Db	165	SCCMSMYMGSCGYTRGWKWSKYSMCKKYCSCTKYCSYTYRYCKMYKYSYKY	106
QY	861	CAAGAGGCTGGAGGCGAGGACCTCACTCACTGGCCAGCAAGGTTGGGGCCCGCAGCAGC	920
Db	105	YCYCYWYSYMYRMKMCNCRSCRSWSMCAVCSTSTSRWMSMYAAKMGWCGSSGYRM	46
QY	921	CCCTCTCTGT	929
Db	45	SKSCWYSK	37
RESULT 11			
ADQ22659			
ID	ADQ22659	standard; DNA; 1813 BP.	
XX			
AC	ADQ22659;		
XX			
DT	26-AUG-2004	(first entry)	
XX			
DE	Human soft tissue sarcoma-upregulated DNA - SEQ ID 5479.		
XX			
KW	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;		
KW	ds.		
XX			
OS	Homo sapiens.		
XX			
FN	WO2004048938-A2.		
XX			
PD	10-JUN-2004.		
XX			
PF	26-NOV-2003; 2003WO-US038193.		
XX			
PR	26-NOV-2002; 2002US-0429739P.		
XX			
PA	(PROT-) PROTEIN DESIGN LABS INC.		
XX			
PI	Aziz N, Ginsburg WM, Zlotnik A;		
XX	WPI; 2004-441208/41.		
DR			

sarcoma.

Example 2; SEQ ID NO 5479; 210pp; English.

The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cyrostatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated DNA of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

Sequence 1813 BP; 410 A; 590 C; 505 G; 290 T; 0 U; 18 Other:

Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue


```
Db      544 ATCCACATCCGAGAG 558      ||| | | |||
RESULT 14
ADQ21633
ID      ADQ21633 standard; DNA; 3613 BP.
XX
XX      AC      ADQ21633;
XX
XX      DT      26-AUG-2004 (first entry)
XX
DE      Human soft tissue sarcoma-upregulated DNA - SEQ ID 4453.
XX
XX      KW      soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX      KW      ds.
XX
XX      OS      Homo sapiens.
XX
XX      PN      WO2004048938-A2.
XX
XX      PD      10-JUN-2004.
XX
XX      PF      26-NOV-2003; 2003WO-US038193.
XX
XX      PR      26-NOV-2002; 2002US-0429739P.
XX
XX      PA      (PROT-) PROTEIN DESIGN LABS INC.
XX
XX      PI      Aziz N, Ginsburg WM, Zlotnik A;
XX
XX      DR      WPI; 2004-441208/41.
XX
XX      PT      Early detection of soft tissue sarcoma comprises determining expression
XX      PT      of a gene in a first soft tissue sample and a normal soft tissue sample
XX      PT      and comparing the gene expression, also useful in treating soft tissue
XX      PT      sarcoma.
XX
XX      PS      Example 2; SEQ ID NO 4453; 210pp; English.
XX
XX      CC      The invention relates to a novel method for detecting soft tissue sarcoma
XX      CC      which comprises obtaining a first soft tissue sample from an individual
XX      CC      and a normal soft tissue sample from the same or different individual,
XX      CC      determining the expression of a gene in both samples and comparing the
XX      CC      expression of the gene in both soft tissue samples, where a higher level
XX      CC      of protein expression in the first soft tissue sample indicates the
XX      CC      presence of soft tissue sarcoma. The method of the invention has
XX      CC      cytosolic applications and may be useful for detecting soft tissue
XX      CC      sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX      CC      acid sequences may be useful in diagnostic and screening applications.
XX      CC      The current sequence is that of a human soft tissue sarcoma-upregulated
XX      CC      DNA of the invention. The current sequence is not shown within the
XX      CC      specification per se but was submitted in CD format by the inventor.
XX
XX      SQ      Sequence 3613 BP; 778 A; 1026 C; 1185 G; 624 T; 0 U; 0 Other;

Query Match      4.2%; Score 46.6; DB 12; Length 3613;
Best Local Similarity 53.8%; Pred. No. 1.8;
Matches 119; Conservative 0; Mismatches 99; Indels 3; Gaps 1;

Qy      6 CAGCCACACTGTGACACTGTCTTGGGAGCTCCCATGATGATGAGAGCGGTGTCTGTG 65
Db      1743 CAGCCAGGGCATCACATGACATGGACAGCACCTCGGGGCCCGGAGCGCCACATCCT 1802

Qy      66 TGGCTACCGCTGGAGTG---AAGGAGGGGCCACAGGCCAGTGGCGGTGTGCCACGA 122
Db      1803 GGGCTACCTGATCGAGAGGCGTAAAGAGGGGAGCAACACCTGGACGGCAGTGAACGACCA 1862

Qy      123 GCTGTGCTTGGACCCCGAGTGTGTGGTGGATGGCTTGGCCCCCGGGGAGACTTACCCTTT 182
Db      1863 GCCGGTGCCTGAGAGGAGTGGACGTGTGGCGGAGCTGTGGCAGGCGCTGTCAATATGAGTT 1922

Qy      183 CCGTGTGGCAGCTGTGGGCCCTGTGGGTGTGTGGGGAACCGG 223

Db      1923 CCGGGTCACAGCTGTGGCTCCCTCAGGTCCCGGAGAGCCTG 1963

RESULT 15
ADQ25448
ID      ADQ25448 standard; DNA; 4176 BP.
XX
XX      AC      ADQ25448;
XX
XX      DT      26-AUG-2004 (first entry)
XX
DE      Human soft tissue sarcoma-upregulated DNA - SEQ ID 8268.
XX
XX      KW      soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX      KW      ds.
XX
XX      OS      Homo sapiens.
XX
XX      PN      WO2004048938-A2.
XX
XX      PD      10-JUN-2004.
XX
XX      PF      26-NOV-2003; 2003WO-US038193.
XX
XX      PR      26-NOV-2002; 2002US-0429739P.
XX
XX      PA      (PROT-) PROTEIN DESIGN LABS INC.
XX
XX      PI      Aziz N, Ginsburg WM, Zlotnik A;
XX
XX      DR      WPI; 2004-441208/41.
XX
XX      PT      Early detection of soft tissue sarcoma comprises determining expression
XX      PT      of a gene in a first soft tissue sample and a normal soft tissue sample
XX      PT      and comparing the gene expression, also useful in treating soft tissue
XX      PT      sarcoma.
XX
XX      PS      Example 2; SEQ ID NO 8268; 210pp; English.
XX
XX      CC      The invention relates to a novel method for detecting soft tissue sarcoma
XX      CC      which comprises obtaining a first soft tissue sample from an individual
XX      CC      and a normal soft tissue sample from the same or different individual,
XX      CC      determining the expression of a gene in both samples and comparing the
XX      CC      expression of the gene in both soft tissue samples, where a higher level
XX      CC      of protein expression in the first soft tissue sample indicates the
XX      CC      presence of soft tissue sarcoma. The method of the invention has
XX      CC      cytosolic applications and may be useful for detecting soft tissue
XX      CC      sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX      CC      acid sequences may be useful in diagnostic and screening applications.
XX      CC      The current sequence is that of a human soft tissue sarcoma-upregulated
XX      CC      DNA of the invention. The current sequence is not shown within the
XX      CC      specification per se but was submitted in CD format by the inventor.
XX
XX      SQ      Sequence 4176 BP; 967 A; 1133 C; 1364 G; 705 T; 0 U; 7 Other;

Query Match      4.2%; Score 46.6; DB 12; Length 4176;
Best Local Similarity 53.8%; Pred. No. 1.8;
Matches 119; Conservative 0; Mismatches 99; Indels 3; Gaps 1;

Qy      6 CAGCCACACTGTGACACTGTCTTGGGAGCTCCCATGATGATGAGAGCGGTGTCTGTG 65
Db      1745 CAGCCAGGGCATCACATGACATGGACAGCACCTCGGGGCCCGGAGCGCCACATCCT 1804

Qy      66 TGGCTACCGCTGGAGTG---AAGGAGGGGCCACAGGCCAGTGGCGGTGTGCCACGA 122
Db      1805 GGGCTACCTGATCGAGAGGCGTAAAGAGGGGAGCAACACCTGGACGGCAGTGAACGACCA 1864

Qy      123 GCTGTGCTTGGACCCCGAGTGTGTGGTGGATGGCTTGGCCCCCGGGGAGACTTACCCTTT 182
Db      1865 GCCGGTGCCTGAGAGGAGTGGACGTGTGGCGGAGCTGTGGCAGGCGCTGTCAATATGAGTT 1924

Qy      183 CCGTGTGGCAGCTGTGGGCCCTGTGGGTGTGTGGGGAACCGG 223
```

Db 1925 CCGGTCACAGCTGTGGCTCCCTCAGGTCCCGGAGGCTG 1965

Search completed: March 20, 2005, 18:55:07
Job time : 644.09 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:15:10 ; Search time 232.766 Seconds
(without alignments)
10198.307 Million cell updates/sec

Title: US-10-077-130-6_COPY_8500_8900

Perfect score: 401
Sequence: 1 gaagaccagtggtggcgcc.....caccagccagccagggagg 401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:.*
1: geneseqn1980s:.*
2: geneseqn1990s:.*
3: geneseqn2000s:.*
4: geneseqn2001as:.*
5: geneseqn2001bs:.*
6: geneseqn2002as:.*
7: geneseqn2002bs:.*
8: geneseqn2003as:.*
9: geneseqn2003bs:.*
10: geneseqn2003cs:.*
11: geneseqn2003ds:.*
12: geneseqn2004as:.*
13: geneseqn2004bs:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	100.0	2534	11 ADM03261	Adm03261 Human cDN
2	401	100.0	7564	8 ABZ24581	ABZ24581 Human cel
3	401	100.0	14061	6 ABV99363	ABV99363 Human NOV
4	401	100.0	14109	6 ABV99362	ABV99362 Human NOV
5	401	100.0	20489	12 ADQ922881	Adq922881 Human sof
6	401	100.0	24120	8 ABX11642	ABx11642 Human ser
7	393	98.0	1005	6 ABN21414	ABn21414 Human ORF
8	60.6	15.1	2559	8 ACC46215	Acc46215 Human dit
9	60.6	15.1	2737	4 AAI59251	AAi59251 Human pol
10	60.6	15.1	3999	4 AAI10337	AAi10337 Human pol
11	60.6	15.1	5378	13 ADR67197	ADr67197 Human bla
12	60.6	15.1	5382	10 ADD14722	ADD14722 Human src
13	60.6	15.1	20565	4 AAK89083	AAk89083 Human dig
14	60.6	15.1	20565	4 AAS28908	AAs28908 Human lmm
15	60.6	15.1	20565	4 AAL03357	AAi03357 Human rep
16	60.6	15.1	20565	10 ADB31749	ADb31749 Human nov
17	59.8	14.9	31595	10 ADF81661	ADf81661 Leukaemia
18	59.8	14.9	81940	4 AAS05390	AAs05390 Human tit
19	59.8	14.9	81940	6 ABK64829	ABk64829 Human ben
20	59.8	14.9	81940	12 ADQ17315	ADq17315 Human sof

21	59.8	14.9	93801	9 ABX13540	ABx13540 Human RGS
22	59.8	14.9	103052	13 ADQ89963	Adq89963 Antagonis
23	59	14.7	2184	4 AAH18256	AaH18256 Human cDN
24	56	14.0	716	6 ABK35726	ABk35726 cDNA seq
25	56	14.0	1266	6 ABK35725	ABk35725 cDNA seq
26	56	14.0	1645	5 AAF24162	AaF24162 Human esc
27	56	14.0	1948	2 ABV74346	ABv74346 Human IL-
28	56	14.0	1949	2 ABV74347	ABv74347 Human IL-
29	56	14.0	2170	6 ABQ54970	ABq54970 Human ova
30	56	14.0	9591	4 AAS28906	AAs28906 Human lmm
31	56	14.0	9591	4 AAL03355	AAi03355 Human rep
32	56	14.0	9591	10 ADB31747	ADb31747 Human nov
33	56	14.0	12415	4 AAS28907	AAs28907 Human lmm
34	56	14.0	12415	4 AAL03356	AAi03356 Human rep
35	56	14.0	12415	10 ADB31748	ADb31748 Human nov
36	55.8	13.9	466	12 ACH89436	ACH89436 Human gen
37	55	13.7	724	4 AAH08421	AaH08421 Human cDN
38	54.8	13.7	4302	13 ACN41312	ACn41312 Human dia
39	54.8	13.7	7328	4 AAK89081	AAk89081 Human dig
40	54.4	13.6	2768	8 ABX71198	ABx71198 Novel hum
41	52.8	13.2	3956	13 ADR08183	ADr08183 Full leng
42	52.6	13.1	572	5 AAS42508	AAs42508 Human cDN
43	52.6	13.1	1251	10 ADC30509	ADc30509 Human nov
44	52.6	13.1	2488	4 AAH18453	AaH18453 Human cDN
45	51.2	12.8	2000	8 ADA71938	ADa71938 Rice gene

ALIGNMENTS

RESULT 1
ADM03261
ID ADM03261 standard; cDNA; 2534 BP.
XX
AC ADM03261;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cDNA of the invention SEQ ID NO:1946.
XX
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EPI347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
(REAS-) RES ASSOC BIOTECHNOLOGY.

Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S; Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I; Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y; WPI; 2003-723558/69.
P-PSDB; ADM05704.

New polynucleotides and polypeptides are useful in gene therapy, for developing a diagnostic marker or medicines for regulating their expression and activity, or as a target of gene therapy.

Claim 1; SEQ ID NO 1946; 305pp; English.
The invention relates to a novel human polynucleotide and the encoded polypeptide. A polynucleotide of the invention may have a use in gene therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful as a primer for synthesizing the polynucleotide or as a probe for detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are useful in gene therapy, for developing a diagnostic marker or medicines

CC for regulating their expression and activity, or as a target of gene
 CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
 CC are useful as pharmaceutical agents. The present sequence represents a
 CC cDNA sequence of the invention.

XX SQ Sequence 2534 BP; 551 A; 679 C; 778 G; 526 T; 0 U; 0 Other;

Query Match 100.0%; Score 401; DB 11; Length 2534;
 Best Local Similarity 100.0%; Pred. No. 1.4e-86;
 Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGACCAAGTGGTGGCGCCAGGGAGGAGCTGGAGCTGGCTGTGAGTGTACAGGGCG 60
 Db |||||
 QY 61 GGAACGCCCTGGCTGAGGACAGGAAGGCCATCCGCAAGACCGCAAGATATGAT 120
 Db |||||
 QY 868 GGAACGCCCTGGCTGAGGACAGGAAGGCCATCCGCAAGACCGCAAGATATGAT 927
 QY 121 GTGGTCTGGAGGGACGATGGCCATGCTGCTATCCGGGGGCTCGCTCAAGGACGCG 180
 Db |||||
 QY 928 GTGGTCTGGAGGGACGATGGCCATGCTGCTATCCGGGGGCTCGCTCAAGGACGCG 987
 QY 181 GCGGAGTACACGTGTGAGTGGAGGCTTCCAGAGACAGCCAGCGCTTCATGTGAAGAA 240
 Db |||||
 QY 988 GCGGAGTACACGTGTGAGTGGAGGCTTCCAGAGACAGCCAGCGCTTCATGTGAAGAA 1047
 QY 241 AAAGCAAACTGCTTTCACAGAGAGTGTACCAATCTGCAAGTGGAGGAGAAAGGCACGCT 300
 Db |||||
 QY 1048 AAAGCAAACTGCTTTCACAGAGAGTGTACCAATCTGCAAGTGGAGGAGAAAGGCACGCT 1107
 QY 301 GTGTTACGTGTGAAGACGAGACACCCCGGGCCACAGTGAAGTGGCGCAAGGGGCTCTTG 360
 Db |||||
 QY 1108 GTGTTACGTGTGAAGACGAGACACCCCGGGCCACAGTGAAGTGGCGCAAGGGGCTCTTG 1167
 QY 361 GAGCTACGGGCTTCAGGGAAGCAGCAGCCAGCCAGCCAGGAGGG 401
 Db |||||
 QY 1168 GAGCTACGGGCTTCAGGGAAGCAGCAGCCAGCCAGCCAGGAGGG 1208

RESULT 2

ABZ24581
 ID ABZ24581 standard; cDNA; 7564 BP.

XX AC ABZ24581;

XX DT 31-MAR-2003 (first entry)

XX DE Human cell adhesion and extracellular matrix protein 4 cDNA.

XX KW Cell adhesion and extracellular matrix protein 4; CADECM-4; human;
 KW anti-HIV; virucide; anti-allergic; anti-inflammatory; anti-naemic;
 KW antiparkinsonian; nootropic; anticonvulsant; antiinfertility;
 KW antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
 KW cytosolic; hepatotropic; dermatological; antidiabetic; nephrotropic;
 KW antigout; thymimetic; neuroprotective; osteopathic; antiarthritis;
 KW antiparasitic; antihelminthic; antipsoriatic; uropathic; ophthalmological;
 KW antirheumatic; haemostatic; antibacterial; protozoacide; fungicide;
 KW synaecological; titin; gene therapy; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 241..7227

XX FT /*tag= a

XX FT /product= "Human CADECM-4"

XX XX WO20028322-A2.

XX XX 07-NOV-2002.

XX XX 01-MAY-2002; 2002WO-US013874.

PR 02-MAY-2001; 2001US-0298290P.
 PR 21-MAY-2001; 2001US-0292468P.
 PR 15-JUN-2001; 2001US-0298616P.
 PR 28-JUN-2001; 2001US-0301672P.
 PR 04-JAN-2002; 2002US-0345008P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding L;
 PI Hillman JL, Baughn MR, Kallick DA, Lee S, Warren BA, Xu Y, Tran UK;
 PI Lal PG, Thornton M, Hafalia AJA, Yao MG, Nguyen DB, Gandhi AR;
 PI Khan FA, Walha NK, Griffin JA, Chinn AM, Elliott VS, Ramkumar J;
 PI Arvizu CS, Forsythe LJ;

XX WPI; 2003-167112/16.
 DR P-PSDB; ABP58227.

XX New human cell adhesion and extracellular matrix proteins, useful for
 diagnosing, treating or preventing autoimmune or inflammatory disorder
 (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
 cancer or hepatitis.

XX Claim 12; Page 168-170; 178pp; English.

XX The present sequence is that of Incyte clone 7326129CB1 cDNA encoding
 human cell adhesion and extracellular matrix protein 4 (CADECM-4). A
 representative cDNA library for the full-length polynucleotide is
 MUSLDR02, constructed from the muscle tissue RNA of a Caucasian adult
 man. Homology data suggest the encoded protein to a titin muscle protein.
 The invention provides CADECM-1 to -11 polypeptides (see ABP58224-34) and
 polynucleotides (see ABZ24578-88), expression vectors, host cells,
 antibodies, agonists and antagonists. These are useful for diagnosing,
 treating or preventing disorders associated with aberrant expression of
 CADECM, particularly cell proliferative disorders (e.g. arteriosclerosis,
 atherosclerosis, cirrhosis, hepatitis, psoriasis, paroxysmal nocturnal
 haemoglobinuria, polycythaemia vera, psoriasis, primary
 thrombocytopaenia or cancer), developmental disorders (e.g. renal
 tubular acidosis, anaemia or mental retardation), neurological disorders
 (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), reproductive
 disorders (e.g. infertility or a disruption in the menstrual cycle), or
 autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma, autoimmune
 thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus,
 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease,
 Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
 osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid
 arthritis, Sjogren's syndrome, uveitis), or viral, bacterial, fungal,
 parasitic, protozoal or helminthic infections

XX SQ Sequence 7564 BP; 1594 A; 1999 C; 2711 G; 1260 T; 0 U; 0 Other;

Query Match 100.0%; Score 401; DB 8; Length 7564;
 Best Local Similarity 100.0%; Pred. No. 1.7e-86;
 Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGACCAAGTGGTGGCGCCAGGGAGGACGTGGAGCTGCGTGTGAGTGTACAGGGCG 60
 Db |||||
 QY 2617 GAAGACCAAGTGGTGGCGCCAGGGAGGACGTGGAGCTGCGTGTGAGTGTACAGGGCG 2676
 QY 61 GGAACGCCCTGGCTGAGGACAGGAAGGCCATCCGCAAGACCGCAAGATATGAT 120
 Db |||||
 QY 2677 GGAACGCCCTGGCTGAGGACAGGAAGGCCATCCGCAAGACCGCAAGATATGAT 2736
 QY 121 GTGGTCTGGAGGGCACGATGGCCATGCTGGTTCATCCCGGGGCTCGCTCAAGGACGCG 180
 Db |||||
 QY 2737 GTGGTCTGGAGGGCACGATGGCCATGCTGGTTCATCCCGGGGCTCGCTCAAGGACGCG 2796
 QY 181 GCGGAGTACAGTGTGAGGTGGAGGCTTCCAAAGACACAGCCAGCCTCCATGTGGAAGAA 240
 Db |||||
 QY 2797 GCGGAGTACAGTGTGAGGTGGAGGCTTCCAAAGACACAGCCAGCCTCCATGTGGAAGAA 2856
 QY 241 AAAGCAAACTGCTTTCACAGAGAGCTGACCAATCTGCAAGGTGGAGGAGAAAGGCACGCT 300
 Db |||||
 QY 2857 AAAGCAAACTGCTTTCACAGAGAGCTGACCAATCTGCAAGGTGGAGGAGAAAGGCACGCT 2916

QY 301 GTGTTACGTCACAGCAGGACCCCGGGCCACAGTACCTGGCGCAAGGGCTCTTG 360
 DB 2917 GTGTTACGTCACAGCAGGACCCCGGGCCACAGTACCTGGCGCAAGGGCTCTTG 2976
 QY 361 GAGCTACGGGCTCAGGGAAGCACCAGCCAGCCAGCCAGGAGGG 401
 DB 2977 GAGCTACGGGCTCAGGGAAGCACCAGCCAGCCAGGAGGG 3017

RESULT 3
 ABV99363
 ID ABV99363 standard; DNA; 14061 BP.
 XX
 AC ABV99363;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Human NOVI3b coding sequence.
 XX
 KW Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS;
 KW antinflammatory; cardiac; haemostatic; neuroprotective; anorectic;
 KW nootropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;
 KW antifertility; cerebroprotective; gene therapy; NOVX; NOV; fertility;
 KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;
 KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
 KW immune disorder; haematopoietic disorder; cardiovascular disorder;
 KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;
 KW metabolic syndrome X; wasting disorder; cell differentiation; gene;
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200272771-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 08-MAR-2002; 2002WO-US007288.
 XX
 PR 08-MAR-2001; 2001US-0274101P.
 PR 08-MAR-2001; 2001US-0274194P.
 PR 08-MAR-2001; 2001US-0274281P.
 PR 08-MAR-2001; 2001US-0274322P.
 PR 09-MAR-2001; 2001US-0274849P.
 PR 12-MAR-2001; 2001US-0275233P.
 PR 13-MAR-2001; 2001US-0275578P.
 PR 13-MAR-2001; 2001US-0275579P.
 PR 13-MAR-2001; 2001US-0275601P.
 PR 14-MAR-2001; 2001US-0276000P.
 PR 16-MAR-2001; 2001US-0276778P.
 PR 19-MAR-2001; 2001US-0276994P.
 PR 20-MAR-2001; 2001US-0277239P.
 PR 20-MAR-2001; 2001US-0277321P.
 PR 20-MAR-2001; 2001US-0277327P.
 PR 20-MAR-2001; 2001US-0277338P.
 PR 21-MAR-2001; 2001US-0277791P.
 PR 22-MAR-2001; 2001US-0277833P.
 PR 23-MAR-2001; 2001US-0278152P.
 PR 26-MAR-2001; 2001US-0278894P.
 PR 27-MAR-2001; 2001US-0278999P.
 PR 27-MAR-2001; 2001US-0279038P.
 PR 28-MAR-2001; 2001US-0279344P.
 PR 30-MAR-2001; 2001US-0279995P.
 PR 30-MAR-2001; 2001US-0280233P.
 PR 02-APR-2001; 2001US-0280802P.
 PR 02-APR-2001; 2001US-0280822P.
 PR 02-APR-2001; 2001US-0280900P.
 PR 04-APR-2001; 2001US-0281194P.
 PR 13-APR-2001; 2001US-0283675P.
 PR 30-APR-2001; 2001US-0287424P.
 PR 02-MAY-2001; 2001US-0288066P.
 PR 03-MAY-2001; 2001US-0288342P.
 PR 03-MAY-2001; 2001US-0288528P.

PR 15-MAY-2001; 2001US-0291190P.
 PR 16-MAY-2001; 2001US-0291099P.
 PR 16-MAY-2001; 2001US-0291240P.
 PR 30-MAY-2001; 2001US-0294485P.
 PR 31-MAY-2001; 2001US-0294889P.
 PR 31-MAY-2001; 2001US-0294899P.
 PR 18-JUN-2001; 2001US-0299027P.
 PR 19-JUN-2001; 2001US-0299303P.
 PR 19-JUN-2001; 2001US-0299310P.
 PR 10-JUL-2001; 2001US-0304354P.
 PR 31-JUL-2001; 2001US-0309198P.
 PR 16-AUG-2001; 2001US-0312903P.
 PR 10-SEP-2001; 2001US-0318462P.
 PR 12-SEP-2001; 2001US-0318770P.
 PR 27-SEP-2001; 2001US-0325430P.
 PR 27-SEP-2001; 2001US-0325681P.
 PR 18-OCT-2001; 2001US-0330380P.
 PR 31-OCT-2001; 2001US-0335301P.
 PR 14-NOV-2001; 2001US-0332172P.
 PR 14-NOV-2001; 2001US-0332271P.
 PR 14-NOV-2001; 2001US-0332272P.
 PR 14-NOV-2001; 2001US-0333184P.
 PR 21-NOV-2001; 2001US-0333272P.
 PR 03-DEC-2001; 2001US-0332094P.
 PR 03-DEC-2001; 2001US-0337426P.
 PR 03-DEC-2001; 2001US-0338092P.
 PR 04-DEC-2001; 2001US-0337185P.
 PR 03-JAN-2002; 2002US-0345705P.
 PR 08-MAR-2002; 2002US-00093463.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
 PI Boldog FL, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Vernet CAM;
 PI Pena CE, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;
 PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
 PI Taupier RJ, Padigaru M, Shenoy SG, Kekuda R, Gusev VY, Pochart PF;
 PI Zhong M;
 XX
 DR WPI; 2002-732824/79.
 DR P-PSDB; ABP70085.
 XX
 PT New NOVX polypeptides and polynucleotides, useful for preventing,
 PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
 PT Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic
 PT disorders, and asthma.
 XX
 PS Claim 16; Page 138-142; 619pp; English.
 XX
 CC The present invention relates to new isolated proteins (NOVX) and their
 CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
 CC any number from 1 to 48. The NOVX proteins and coding sequences are
 CC useful in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, preferably a NOVX-associated disorder.
 CC The NOVX coding sequences and proteins are useful for treating,
 CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
 CC obesity, infectious disease, anorexia, cancer-associated cachexia,
 CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
 CC disease, immune disorders, haematopoietic disorders, cardiovascular
 CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
 CC disturbances associated with obesity, metabolic syndrome X or wasting
 CC disorders associated with chronic diseases or various cancers. The NOVX
 CC coding sequences and proteins may also be used as targets for the
 CC identification of small molecules that modulate or inhibit e.g.
 CC neurogenesis, cell differentiation, cell proliferation, haematopoiesis,
 CC wound healing and angiogenesis, in gene therapy, in generation of
 CC antibodies that bind immunospecifically to NOVX substances for use in
 CC therapeutic or diagnostic methods
 XX
 SQ Sequence 14061 BP; 2854 A; 4029 C; 4839 G; 2339 T; 0 U; 0 Other;
 Query Match 100.0%; Score 401; DB 6; Length 14061;
 Best Local Similarity 100.0%; Pred. No. 2e-86;

Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	GAAGACAGTGGGTGGCGCCAGGGGAGGAGCTGCAGCTGGCTGTGAGCTGTGACGGGGG 60	
Db	732	GAAGACAGTGGGTGGCGCCAGGGGAGGAGCTGGCTGTGAGCTGTGACGGGGG 791	
Qy	61	GGAAACCCCGTGCACCTGGCTGGAAGGACAGGAAGGCCATCCGCAAGAGCCAGAAGTATGAT 120	
Db	792	GGAAACCCCGTGCACCTGGCTGGAAGGACAGGAAGGCCATCCGCAAGAGCCAGAAGTATGAT 851	
Qy	121	GTGGTCTGCGAGGGGACGATGGCCATGCTGCTCATCCGGGGGCTCGCTCAAGGACGG 180	
Db	852	GTGGTCTGCGAGGGGACGATGGCCATGCTGCTCATCCGGGGGCTCGCTCAAGGACGG 911	
Qy	181	GGCGAGTACACGTGTGAGGTGGAGGCTTCCAGAGCACAGCCAGCCTCCATGTGGAAGAA 240	
Db	912	GGCGAGTACACGTGTGAGGTGGAGGCTTCCAGAGCACAGCCAGCCTCCATGTGGAAGAA 971	
Qy	241	AAAGCAAACTGTTTACAGAGGAGCTGACCAATCTGCAAGTGGAGGAGAAAGCCACAGCT 300	
Db	972	AAAGCAAACTGTTTACAGAGGAGCTGACCAATCTGCAAGTGGAGGAGAAAGCCACAGCT 1031	
Qy	301	GTGTTTCACTGCAAGACGAGCACCCCGGGGCCACAGTGACCTGGCGCAAGGGCTCTTTG 360	
Db	1032	GTGTTTCACTGCAAGACGAGCACCCCGGGGCCACAGTGACCTGGCGCAAGGGCTCTTTG 1091	
Qy	361	GAGCTACGGGCTTCAGGGGAAGCACCCAGCCAGCCAGGAGGG 401	
Db	1092	GAGCTACGGGCTTCAGGGGAAGCACCCAGCCAGCCAGGAGGG 1132	

RESULT 4

ID ABV99362 standard; DNA; 14109 BP.

AC ABV99362;

DT 27-JAN-2003 (first entry)

DE Human NOVI3a coding sequence.

KW Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS;
KW antiinflammatory; cardiant; haemostatic; neuroprotective; anorectic;
KW neotropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;
KW antifertility; cerebroprotective; gene therapy; NOVI; NOV; fertility;
KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;
KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; cardiovascular disorder;
KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;
KW metabolic syndrome X; wasting disorder; cell differentiation; gene;
KW cell proliferation; haematopoiesis; wound healing; angiogenesis; ds.

OS Homo sapiens.

PN WO200272771-A2.

PD 19-SEP-2002.

PF 08-MAR-2002; 2002WO-US007288.

PR 08-MAR-2001; 2001US-0274101P.

PR 08-MAR-2001; 2001US-0274194P.

PR 08-MAR-2001; 2001US-0274281P.

PR 09-MAR-2001; 2001US-0274322P.

PR 12-MAR-2001; 2001US-0274849P.

PR 13-MAR-2001; 2001US-0275235P.

PR 13-MAR-2001; 2001US-0275579P.

PR 14-MAR-2001; 2001US-0275601P.

PR 16-MAR-2001; 2001US-0276000P.

PR 19-MAR-2001; 2001US-0276994P.

PR 20-MAR-2001; 2001US-0277239P.

PR 20-MAR-2001; 2001US-0277321P.
PR 20-MAR-2001; 2001US-0277327P.
PR 20-MAR-2001; 2001US-0277338P.
PR 21-MAR-2001; 2001US-0277791P.
PR 22-MAR-2001; 2001US-0277833P.
PR 23-MAR-2001; 2001US-0278152P.
PR 26-MAR-2001; 2001US-0278894P.
PR 27-MAR-2001; 2001US-0278999P.
PR 27-MAR-2001; 2001US-0279036P.
PR 28-MAR-2001; 2001US-0279344P.
PR 30-MAR-2001; 2001US-0279995P.
PR 30-MAR-2001; 2001US-0280233P.
PR 02-APR-2001; 2001US-0280802P.
PR 02-APR-2001; 2001US-0280822P.
PR 04-APR-2001; 2001US-0280900P.
PR 13-APR-2001; 2001US-0281194P.
PR 30-APR-2001; 2001US-0283675P.
PR 03-MAY-2001; 2001US-0288066P.
PR 03-MAY-2001; 2001US-0288342P.
PR 03-MAY-2001; 2001US-0288528P.
PR 15-MAY-2001; 2001US-0291190P.
PR 16-MAY-2001; 2001US-0291099P.
PR 16-MAY-2001; 2001US-0291240P.
PR 30-MAY-2001; 2001US-0294485P.
PR 31-MAY-2001; 2001US-0294889P.
PR 31-MAY-2001; 2001US-0294899P.
PR 18-JUN-2001; 2001US-0299027P.
PR 19-JUN-2001; 2001US-0299303P.
PR 19-JUN-2001; 2001US-0299310P.
PR 10-JUL-2001; 2001US-0304354P.
PR 31-JUL-2001; 2001US-0309198P.
PR 16-AUG-2001; 2001US-0312903P.
PR 10-SEP-2001; 2001US-0318462P.
PR 12-SEP-2001; 2001US-0318770P.
PR 27-SEP-2001; 2001US-0325430P.
PR 27-SEP-2001; 2001US-0325681P.
PR 18-OCT-2001; 2001US-0330380P.
PR 31-OCT-2001; 2001US-0335301P.
PR 14-NOV-2001; 2001US-0332172P.
PR 14-NOV-2001; 2001US-0332271P.
PR 14-NOV-2001; 2001US-0332272P.
PR 14-NOV-2001; 2001US-0333184P.
PR 21-NOV-2001; 2001US-0333272P.
PR 03-DEC-2001; 2001US-0332094P.
PR 03-DEC-2001; 2001US-0337426P.
PR 04-DEC-2001; 2001US-0338092P.
PR 03-JAN-2002; 2002US-0345705P.
PR 08-MAR-2002; 2002US-00093463.

(CURA-) CURAGEN CORP.

Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
Boidog FL, Li L, Zehrhusen BD, Tchernev VT, Gangolli EA, Vernet
Pena CE, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK,
Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
Taupier RJ, Padigaru M, Shenoy SS, Kekuda R, Gusev VY, Pochart PF;
Zhong M;

WPI; 2002-732824/79.
P-PSDB; ABP70084.

New NOVI polypeptides and polynucleotides, useful for preventing,
diagnosing or treating NOVI-associated disorders e.g. diabetes, cancer,
Alzheimer's disease, dyslipidemia, obesity, immune or hematopoietic
disorders, and asthma.

Claim 16; Page 133-137; 619pp; English.

The present invention relates to new isolated proteins (NOVI) and their
coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
any number from 1 to 48. The NOVI proteins and coding sequences are

CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, preferably a NOVX-associated disorder.
CC The NOVX coding sequences and proteins are useful for treating,
CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
CC obesity, infectious diseases, anorexia, cancer-associated cachexia,
CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
CC disease, immune disorders, haematopoietic disorders, cardiovascular
CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
CC disturbances associated with obesity, metabolic syndrome X or wasting
CC disorders associated with chronic diseases or various cancers. The NOVX
CC coding sequences and proteins may also be used as targets for the
CC identification of small molecules that modulate or inhibit e.g.
CC neurogenesis, cell differentiation, cell proliferation, haematopoiesis,
CC wound healing and angiogenesis, in gene therapy, in generations for use in
CC antibodies that bind immunospecifically to NOVX substances or in
CC therapeutic or diagnostic methods
XX
SQ Sequence 14109 BP; 2862 A; 4045 C; 4854 G; 2348 T; 0 U; 0 Other;
Query Match 100.0%; Score 401; DB 6; Length 14109;
Best Local Similarity 100.0%; Pred. No. 2e-86;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAAGACCACTGGTGGCGCCAGGGAGGACGTGGAGCTGGCTGTGAGCTGTACGGGCG 60
Db 732 GAAGACCACTGGTGGCGCCAGGGAGGACGTGGAGCTGGCTGTGAGCTGTACGGGCG 791
Qy 61 GGAAGCCCGTGCACCTGGCTGAAGACACAGGAGGCCATCCGCAAGAGCCAGAAGTATGAT 120
Db 792 GGAAGCCCGTGCACCTGGCTGAAGACACAGGAGGCCATCCGCAAGAGCCAGAAGTATGAT 851
Qy 121 GTGCTCTCGAGGGACACATGGCCATGTGTCTATCCCGGGGCTCGCTCAAGACCGG 180
Db 852 GTGCTCTCGAGGGACACATGGCCATGTGTCTATCCCGGGGCTCGCTCAAGACCGG 911
Qy 181 GCGAGTACACGTGTGAGTGGAGGCTTCCAGAGCACAGCCAGCTCCATGTGGAAGAA 240
Db 912 GCGAGTACACGTGTGAGTGGAGGCTTCCAGAGCACAGCCAGCTCCATGTGGAAGAA 971
Qy 241 AAAGCAAACTGCTTCCACAGAGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCT 300
Db 972 AAAGCAAACTGCTTCCACAGAGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCT 1031
Qy 301 GTGTTACGTGCAAGACGAGACACCCCGGGCCACAGTGACCTGGCGCAAGGGCTCTTG 360
Db 1032 GTGTTACGTGCAAGACGAGACACCCCGGGCCACAGTGACCTGGCGCAAGGGCTCTTG 1091
Qy 361 GAGCTACGGGCTCAGGGAAGCACCCAGCCAGCCAGCCAGGAGGG 401
Db 1092 GAGCTACGGGCTCAGGGAAGCACCCAGCCAGCCAGCCAGGAGGG 1132
RESULT 5
ADQ22881
ID ADQ22881 standard; DNA; 20489 BP.
XX
AC ADQ22881;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5701.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX ds.
XX Homo sapiens.
XX
FN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX

PR 26-NOV-2002; 2002US-0429739P.
XX (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 5701; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytosstatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 20489 BP; 3917 A; 5815 C; 6983 G; 3225 T; 0 U; 549 Other;
Query Match 100.0%; Score 401; DB 12; Length 20489;
Best Local Similarity 100.0%; Pred. No. 2.1e-86;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAAGACCACTGGTGGCGCCAGGGAGGACGTGGAGCTGGCTGTGAGCTGTACGGGCG 60
Db 8571 GAAGACCACTGGTGGCGCCAGGGAGGACGTGGAGCTGGCTGTGAGCTGTACGGGCG 8630
Qy 61 GGAAGCCCGTGCACCTGGCTGAAGGACACAGGAGGCCATCCGCAAGAGCCAGAAGTATGAT 120
Db 8631 GGAAGCCCGTGCACCTGGCTGAAGGACACAGGAGGCCATCCGCAAGAGCCAGAAGTATGAT 8690
Qy 121 GTGCTCTCGAGGGACACATGGCCATGTGTCTATCCCGGGGCTCGCTCAAGACCGG 180
Db 8691 GTGCTCTCGAGGGACACATGGCCATGTGTCTATCCCGGGGCTCGCTCAAGACCGG 8750
Qy 181 GCGAGTACACGTGTGAGTGGAGGCTTCCAGAGCACAGCCAGCTCCATGTGGAAGAA 240
Db 8751 GCGAGTACACGTGTGAGTGGAGGCTTCCAGAGCACAGCCAGCTCCATGTGGAAGAA 8810
Qy 241 AAAGCAAACTGCTTCCACAGAGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCT 300
Db 8811 AAAGCAAACTGCTTCCACAGAGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCT 8870
Qy 301 GTGTTACGTGCAAGACGAGACACCCCGGGCCACAGTGACCTGGCGCAAGGGCTCTTG 360
Db 8871 GTGTTACGTGCAAGACGAGACACCCCGGGCCACAGTGACCTGGCGCAAGGGCTCTTG 8930
Qy 361 GAGCTACGGGCTCAGGGAAGCACCCAGCCAGCCAGCCAGGAGGG 401
Db 8931 GAGCTACGGGCTCAGGGAAGCACCCAGCCAGCCAGCCAGGAGGG 8971
RESULT 6
ABX11642
ID ABX11642 standard; cDNA; 24120 BP.
XX
AC ABX11642;
XX
DT 09-MAY-2003 (first entry)
XX
DE Human serine/threonine or protein kinase 12599, cDNA.

XX Human; ss; gene; serine/threonine kinase; protein kinase; 12599;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 5'UTR 1..71
FT /*tag= a
FT 72..23978
FT /*tag= b
FT /product= "Kinase 12599"
FT /note= "This CDS is specifically claimed in claim 2"
FT 23979..24120
FT /*tag= c
XX
XX US2002168742-A1.
XX
XX 14-NOV-2002.
XX
XX 15-FEB-2002; 2002US-00077130.
XX
XX 15-FEB-2001; 2001US-0269201P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Kapeller-Libermann R, Acton SL;
XX
XX WPI; 2003-298729/29.
XX P-PSDB; ABG76187.
XX
XX Novel isolated human protein kinase, designated 59079 or 12599
XX polypeptide, useful as diagnostic and therapeutic agents for preventing
XX cardiovascular diseases, proliferative disorders, and protein kinase
XX disorders.
XX
XX Claim 2; Page 58-84; 119pp; English.
XX
XX The invention relates to an isolated human serine/threonine or protein
XX kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
XX comprising at least 85% identity to the nucleic acids appearing as
XX ABX11641 and ABX11642 or their complement, a naturally occurring variant
XX of the kinases or their fragments. Also included are a non-human host
XX cell containing the nucleic acids, an antibody specific for the proteins,
XX identifying a compound which binds to the kinase (by contacting the
XX kinase or a cell expressing the kinase with a test compound and
XX determining whether the kinase binds to the test compound) and modulating
XX the activity of kinase using the identified compound. The kinases and
XX their encoding nucleic acids are useful as diagnostic and therapeutic
XX agents for preventing a disease or condition associated with an aberrant
XX or unwanted 59079 or 12599 activity in a subject, including
XX cardiovascular diseases such as heart failure, and myocardial infarction;
XX disorders involving blood vessels such as atherosclerosis, and Kaposi's
XX sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,
XX Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders
XX such as cancer; and protein kinase disorders such as autoimmune
XX disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,
XX rheumatoid arthritis, and multiple sclerosis (many examples of diseases
XX and disorders are included in the specification). The kinases, their
XX encoding nucleic acids and antibodies are useful in screening assays,
XX detection assays (e.g. forensic biology), and predictive medicine (e.g.
XX diagnostic assays, prognostic assays, and monitoring clinical trials and
XX pharmacogenomics). The kinases and their encoding nucleic acids are
XX useful as query sequences to perform a search against public databases to
XX identify other family members or related sequences. The present sequence
XX encodes the kinase 12599

SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;
Query Match 100.0%; Score 401; DB 8; Length 24120;
Best Local Similarity 100.0%; Pred. No. 2.2e-86;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAAGACCAAGTGGTGGCGCCAGGGAGACGTGAGCTGCGCTGTGAGCTTCAACGGCG 60
Db 8571 GAAGACCAAGTGGTGGCGCCAGGGAGACGTGAGCTGCGCTGTGAGCTTCAACGGCG 8630
Qy 61 GGAAGCCCGTGCACCTGGCTGAAGGACAGGAGGAGCCATCCGACAGAGCCAGAGTATGAT 120
Db 8631 GGAAGCCCGTGCACCTGGCTGAAGGACAGGAGGAGCCATCCGACAGAGCCAGAGTATGAT 8690
Qy 121 GTGCTCTGCGAGGACGATGGCCATGCTGCTCATCCGCGGGGCTCGCTCAAGGACGCG 180
Db 8691 GTGCTCTGCGAGGACGATGGCCATGCTGCTCATCCGCGGGGCTCGCTCAAGGACGCG 8750
Qy 181 GGCAGTACACGTGTGAGTGGAGGCTTCCAAGAGCACAGCCAGCTCCATGTGGAAGAA 240
Db 8751 GGCAGTACACGTGTGAGTGGAGGCTTCCAAGAGCACAGCCAGCTCCATGTGGAAGAA 8810
Qy 241 AAGCAAACTGCTTCACAGAGGAGCTACCAATCTGAGGTGGAGGAGGAGGAGGAGGAGG 300
Db 8811 AAGCAAACTGCTTCACAGAGGAGCTACCAATCTGAGGTGGAGGAGGAGGAGGAGGAGG 8870
Qy 301 GTGTTACGTGCAAGACGAGCACCCCGCGCCACAGTGCACCTCGCGCAAGGGGCTCTTG 360
Db 8871 GTGTTACGTGCAAGACGAGCACCCCGCGCCACAGTGCACCTCGCGCAAGGGGCTCTTG 8930
Qy 361 GAGTACGGGCTCAGGGAAGCACCCAGCCAGCCAGCCAGGAGGG 401
Db 8931 GAGTACGGGCTCAGGGAAGCACCCAGCCAGCCAGGAGGG 8971
RESULT 7
ABN21414
ID ABN21414 standard; cDNA; 1005 BP.
XX AC ABN21414;
XX
XX 24-JUN-2002 (first entry)
XX
XX Human ORFX polynucleotide sequence SEQ ID NO:11305.
XX
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200192523-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US010836.
XX
XX 30-MAY-2000; 2000US-0206132P.
XX 29-AUG-2000; 2000US-0228716P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach MD;
XX WPI; 2002-106308/14.
XX P-PSDB; ABP05662.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders.
PS Disclosure; SEQ ID NO 11305; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABRN15762 to ABRN27252 encode the human ORFX
CC proteins given in ABRN0010 to ABRN1500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC syndrome in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1005 BP; 209 A; 303 C; 340 G; 152 T; 0 U; 1 Other;

Query Match 98.0%; Score 393; DB 6; Length 1005;
Best Local Similarity 98.8%; Pred. No. 9.5e-85;
Matches 396; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GAAGACCACTGGTGGCCGCGGAGGAGCGTGGAGCTGGCTGTGAGCTGTACGCGCG 60
Db 358 GAAGACCACTGGTGGCCGCGGAGGAGCGTGGAGCTGGCTGTGAGCTGTACGCGCG 417

QY 61 GGAAGCCGCGTGGCTGAAGGACAGGAGGCGCATCCGCAAGAGCCAGAAAGTATGAT 120
Db 418 GGAAGCCGCGTGGCTGAAGGACAGGAGGCGCATCCGCAAGAGCCAGAAAGTATGAT 477

QY 121 GTGCTCTCGAGGCGCAGATGCCATGCTGTGTCATCCGCGGGGCTCGCTCAAGACGCG 180
Db 478 GTGCTCTCGAGGCGCAGATGCCATGCTGTGTCATCCGCGGGGCTCGCTCAAGACGCG 537

QY 181 GCGGAGTACAGTGTGAGGTGGAGGCTTCCAGAGCAGACGCGCATCCATGTGGAAGAA 240
Db 538 GCGGAGTACAGTGTGAGGTGGAGGCTTCCAGAGCAGACGCGCATCCATGTGGAAGAA 597

QY 241 AAAGCAAACTGTTCCACAGAGAGCTGACCAATCTGCAGGTGGAGAGAAAGGCACAGCT 300
Db 598 AAAGCAAACTGTTCCACAGAGAGCTGACCAATCTGCAGGTGGAGAGAAAGGCACAGCT 657

QY 301 GTGTTACAGTGAAGACGAGACACCCCGCGGCCACAGTGAAGTGTGCGCAAGGCTCTTG 360
Db 658 GTGTTACAGTGAAGACGAGACACCCCGCGGCCACAGTGAAGTGTGCGCAAGGCTCTTG 717

QY 361 GAGCTACGGGCTTCAGGAGAGCAGCCAGCCAGCCAGGAGGG 401
Db 718 GAGCTACGGGCTTCAGGAGAGCAGCCAGCCAGGAGGG 758

RESULT 8
ACC46215
ID ACC46215 standard; cDNA; 2559 BP.
XX
AC ACC46215;
XX
DT 02-JUN-2003 (first entry)
XX

DE Human dithp intracellular signalling protein-encoding cDNA.
XX
KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW intracellular signalling; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200297031-A2.
XX
PD 05-DEC-2002.
XX
PF 27-MAR-2002; 2002WO-US010056.
XX
PR 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300011P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Ameshey SR;
PI Doughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Garstin BH;
PI Peralta C, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
XX WPI; 2003-129518/12.
DR P-PSDB; ABR41273.
XX
PT Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
PS Claim 2; SEQ ID NO 136; 591pp; English.
XX
CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a dithp cDNA encoding a DITHP protein
CC which has intracellular signalling activity. Note: The sequence data for

CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2559 BP; 548 A; 766 C; 707 G; 538 T; 0 U; 0 Other;
Query Match 15.1%; Score 60.6; DB 8; Length 2559;
Best Local Similarity 54.8%; Pred. No. 8.6e-05;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
Qy 10 TGGGTGGCGCCAGGAGACGTGAGCTGCGCTGTGAGCTGTACGCGGCGGACGCC 69
Db 568 TGTGTGGCCCCCGGAGCCAGTGTGCTGAGCTGTGAACTGTCGCCGGCTGGCGCCCC 627
Qy 70 GTGCACCTGGCTGAAGGACAGGAGCCATCCGCAAGAGCCAGAGTATGATGTGCTGTC 129
Db 628 GTGGTCTGGAGCCACAATGGGAGGCCCTGTGCAGGAGGGCGGAGGCTCCATGCC 687
Qy 130 GAGGACACGATGGCCATGCTGTGCTATCCGCGGGGCTGCTCAAGGAGCGCGGCGAGTAC 189
Db 688 GAGGCGCCCCCGCGAGTCTTGTGCTATCCAGGCTGCAGGCCCCAGCCATGTCAGGGCTCTAC 747
Qy 190 ACCTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCCCTC 228
Db 748 ACCTGCCAGTCTGGAGAGCCCCCGGAGGCCCAAGCCCTC 786
RESULT 9
AAI59251
ID AAI59251 standard; cdna; 2737 BP.
AC AAI59251;
XX
DT 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 1454.
XX
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
FA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR P-ESDB; AAM40095.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.

XX Claim 1; SEQ ID NO 1454; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nontropic, and the
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 2737 BP; 484 A; 840 C; 899 G; 514 T; 0 U; 0 Other;
Query Match 15.1%; Score 60.6; DB 4; Length 2737;
Best Local Similarity 54.8%; Pred. No. 8.8e-05;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
Qy 10 TGGGTGGCGCCAGGAGGACGTGAGCTGCGCTGTGAGCTGTACGCGGCGGAAACGCC 69
Db 1664 TGTGTGGCCCCCGGAGCCAGTGTGCTGAGCTGTGAACCTGTCCCGGCTGGCGCCCC 1723
Qy 70 GTGCACCTGGCTGAAGGACAGGAGCCATCCGCAAGAGCCAGAGTATGATGTGCTGTC 129
Db 1724 GTGGTCTGGAGCCACAATGGGAGGCCCTGTGCAGGAGGGCGGAGCTCCATGCC 1783
Qy 130 GAGGACACGATGGCCATGCTGTGCTATCCGCGGGGCTGCTCAAGGAGCGCGGCGAGTAC 189
Db 1784 GAGGCGCCCCCGCGAGTCTTGTGCTATCCAGGCTGCAGGCCCCAGCCATGAGGGCTCTAC 1843
Qy 190 ACCTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCCCTC 228
Db 1844 ACCTGCCAGTCTGGAGAGGCCCGCGGAGGCCCAAGCCCTC 1882
RESULT 10
AAI61037/c
ID AAI61037 standard; cdna; 3999 BP.
XX
AC AAI61037;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5026.
XX
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
FA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR P-ESDB; AAM40095.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.


```
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSEQ-) HYSEQ INC.
XX
XX Tang YT, Liu C, Auendi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR P-PSDB; AAM41881.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Claim 1; SEQ ID NO 5026; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 3999 BP; 726 A; 1332 C; 1238 G; 703 T; 0 U; 0 Other;

Query Match 15.1%; Score 60.6; DB 4; Length 3999;
Best Local Similarity 54.8%; Pred. No. 9.5e-05;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 TGGGTGGCGCCAGGAGGAGCGTGGAGCTGGCGCTGTGAGCTGTGACGGCGGGAACGCC 69
Db 2612 TGTGTGGCCCTGGGGAGCCAGTGTGTGCTGAGCTGTGACTGTCCGGGCTGGCGCCCC 2553
QY 70 GTGCACCTGGCTGAAGGACGAGGAGCCATCCGCAAGAGCCAGCAAGTATGATGTGCTGTC 129
Db 2552 GTGGTCTGGAGCCACATATGGGAGGCCCTGTGAGGAGGCGAGGGCTTAGAGCTCCATGCC 2493
QY 130 GAGGCGACGATGGCCATCTGTGCTATCCGCGGGGCTGCTCAAGAGCGCGGGCGAGTAC 189
Db 2492 GAGGCGCCCCCGCGAGTCTCTGCAATCCAGGCTGCAGGCCCAAGCCCATGCAGGGCTCTAC 2433
QY 190 ACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCCCTC 228
Db 2432 ACCTGCCAGTCTGGAGAGCCCCCGGAGGCCCAAGCCCTC 2394

RESULT 11
ADR67197
ID ADR67197 standard; DNA; 5378 BP.
XX
XX ADR67197;
XX
XX 02-DEC-2004 (first entry)
XX
XX Human bladder cancer associated nucleotide sequence.
XX
XX bladder cancer tissue; bladder cancer; cytostatic; gene; ds.
XX
XX Homo sapiens.
XX
XX WO2004076613-A2.
XX
```

```
PD 10-SEP-2004.
XX
XX 24-FEB-2004; 2004WO-DE000364.
XX
XX 26-FEB-2003; 2003DE-01009729.
XX
XX (HERR/) HERR A.
PA (HINZ/) HINZMANN B.
PA (DAHL/) DAHL E.
PA (STAU/) STAUB E.
PA (PILA/) PILARSKY C.
PA (SPEC/) SPECHT T.
XX
XX Herr A, Hinzmann B, Dahl E, Staub E, Pilarsky C, Specht T;
XX
XX WPI; 2004-653385/63.
XX
XX New nucleic acids, and encoded proteins, from bladder cancer tissue,
PT useful for diagnosis, treatment and in screening for specific binding
PT agents.
XX
XX Claim 1; Fig 3; 112pp; German.
XX
XX The present invention describes nucleic acids (I) associated with bladder
CC cancer tissue. Also described: (I) peptides and proteins (II) containing
CC an amino acid sequence encoded by (I); (2) a method for diagnosing
CC bladder cancer (BC), or monitoring its progression, that uses (I), (II)
CC or agents (Z) that inhibit, or bind to, (I) or (II), labelled with a
CC reporter; and (3) a method for treating BC that uses (I), (II) or (Z).
CC (I) and (II) have cytostatic activity. (I) and (II) can be used to detect
CC and monitor progression of bladder cancer (BC), or the risk of
CC developing it; to screen for specific binding agents (Z), and to treat
CC BC. (Z) are also useful as diagnostic and therapeutic agents. The present
CC sequence represents a human nucleotide sequence associated with bladder
CC cancer, which is used in the exemplification of the present invention.
XX
SQ Sequence 5378 BP; 940 A; 1696 C; 1799 G; 943 T; 0 U; 0 Other;

Query Match 15.1%; Score 60.6; DB 13; Length 5378;
Best Local Similarity 54.8%; Pred. No. 0.0001;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 TGGGTGGCGCCAGGAGGAGCGTGGAGCTGGCGCTGTGAGCTGTGACGGCGGGAACGCC 69
Db 2782 TGTGTGGCCCTGGGGAGCCAGTGTGTGCTGAGCTGTGAACCTGTCGGGCTGGCGCCCC 2841
QY 70 GTGCACCTGGCTGAAGGACGAGGAGCCATCCGCAAGAGCCAGCAAGTATGATGTGCTGTC 129
Db 2842 GTGTCTGGAGCCACATGGGAGGCCCTGTGAGGAGGCGAGGGCTAGAGCTCCATGCC 2901
QY 130 GAGGCGACGATGGCCATGCTGCTCATCCGCGGGGCTCTCGCTCAAGAGCGCGGGCGAGTAC 189
Db 2902 GAGGCGCCCCCGCGAGTCTCTGCAATCCAGCTGCAGGCCCAAGCCCATGCAGGGCTCTAC 2961
QY 190 ACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCCCTC 228
Db 2962 ACCTGCCAGTCTGGAGAGCCCCCGGAGGCCCAAGCCCTC 3000

RESULT 12
ADD14722
ID ADD14722 standard; cDNA; 5382 BP.
XX
XX AC ADD14722;
XX
XX 01-JAN-2004 (first entry)
XX
XX Human src biomarker polynucleotide SEQ ID NO:116.
XX
XX predictor set; protein tyrosine kinase activity modulator;
XX protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
XX gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
XX
```

OS Homo sapiens.
XX WO2003062395-A2.
XX PD 31-JUL-2003.
XX PF 17-JAN-2003; 2003WO-US001981.
XX PR 18-JAN-2002; 2002US-0350061P.
XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX PI Huang F, Fairchild CR, Lee FY, Shaw P;
XX DR WPI; 2003-636735/60.
XX DR P-PSDB; ADD14123.
XX PT New polynucleotides and polypeptides for predicting the activity of
PT compounds that interact with protein tyrosine kinases and/or protein
PT tyrosine kinase pathways.
XX PS Claim 2; SEQ ID NO 116; 139pp; English.
XX CC The present invention describes a predictor set comprising a plurality of
CC polynucleotides or polypeptides whose expression pattern is predictive of
CC the response of cells to treatment with a compound that modulates protein
CC tyrosine kinase activity or treatment of the protein tyrosine kinase
CC pathway. Also described: (1) predicting whether a compound is capable of
CC modulating the activity of cells, comprising obtaining a sample of cells,
CC determining whether the cells express a plurality of markers, and
CC correlating the expression of the markers to the compound's ability to
CC modulate the activity of the cells; (2) a plurality of cell lines for
CC identifying polynucleotides and polypeptides whose expression levels
CC correlate with compound sensitivity or resistance of cells associated
CC with a disease state; and (3) identifying polynucleotides and
CC polypeptides that predict compound sensitivity or resistance of cells
CC associated with a disease state, comprising subjecting the plurality of
CC cell lines to one or more compounds, analysing the expression pattern of
CC a microarray of polynucleotides or polypeptides, and selecting
CC polynucleotides or polypeptides that predict the sensitivity or
CC resistance of cells associated with a disease state by using the
CC expression pattern of the microarray. The polynucleotides and
CC polypeptides have cytostatic activities, and can be used in gene therapy.
CC The polynucleotides and polypeptides are useful in predicting the
CC activity of compounds that interact with protein tyrosine kinases and/or
CC protein tyrosine kinase pathways. These may be used in determining drug
CC sensitivity in patients to allow the development of individualized
CC genetic profiles which aid in treating diseases and disorders (e.g.
CC cancer) based on patient response at a molecular level. The present
CC sequence is used in the exemplification of the present invention.
XX SQ Sequence 5382 BP; 943 A; 1699 C; 1798 G; 942 T; 0 U; 0 Other;
Query Match 15.1%; Score 60.6; DB 10; Length 5382;
Best Local Similarity 54.8%; Pred. No. 0.0001;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
Qy 10 TGGGTGGCGCCAGGGAGGACGTGGAGCTGGCTGTGTGAGCTGTCTACGGGGGGAACGCCCC 69
Db TGTGTGGCCCCCTGGGGAGCCAGTGGTGTGAGCTGTGAACTGTCCGGGCTGGCGCCCC 2842
Qy 70 GTGCACCTGGCTGAGGACAGGAGCCATCCGGAAGAGCCAGATGATGATGTGCTGC 129
Db GTGTCTGGAGCCCAATGGGAGGCCCGGTGAGGAGGGCGAGGGCCCTAGAGCTCCATGCC 2902
Qy 130 GAGGGCACGATGGCCATGCTGGTCTCCGGGGGCCCTCGTCAAGGACGCGGGCGAGTAC 189
Db GAGGGCCCCCGCCAGGCTCTCTGATCCAGGCTGAGGCGCCAGCCCATGAGGCTCTAC 2962
Qy 190 ACGTGTGAGGTGGAGGCTTCCAGAGCACACGACCCCTC 228
Db ACCTGCCAGTCTGGAGAGGCCCGCGGAGCCCAAGCCCTC 3001

RESULT 13
AAK89083
ID AAK89083 standard; DNA; 20565 BP.
XX AC AAK89083;
XX DT 05-NOV-2001 (first entry)
XX XX Human digestive system antigen genomic sequence SEQ ID NO: 2659.
DE Human; digestive system antigen; Gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX OS Homo sapiens.
XX PN WO200155314-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001324.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-MAR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214986P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.

```
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.

PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-502630/55.
DR
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
XX Disclosure; SEQ ID NO 2659; 986bp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 20565 BP; 4236 A; 6155 C; 5760 G; 4414 T; 0 U; 0 Other;

Query Match 15.1%; Score 60.6; DB 4; Length 20565;
Best Local Similarity 54.8%; Pred. No. 0.00013;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 TGGGTGGCGCCAGGGGAGGACGTGGAGCTGCGCTGTGAGCTGTGTCACGGCGGGGACGCC 69
Db 13230 TGTGTGCCCCCTGGGGAGCCAGTGTGCTGAGCTGTGAACCTGTCGGGCTGGCGCCCC 13289

QY 70 GTGCACTGGCTGAAGGACAGAGGCCATCCGAGAGCCAGAGATGATGTGTGCTTGC 129
Db 13290 GTGCTCTGGAGCCCAATGGAGGCCCGCTGCAGAGGCGGAGGCTAGAGCTCCATGCC 13349

QY 130 GAGGGCAGATGGCCATCTGCTCATCCGGGGGCTCGCTCAAGGACGCGGGCGAGTAC 189
Db 13350 GAGGGCCCCCGCCGAGTCTCTGCTCATCAGGCTCAGGCCCAGCCCATGCAGGGGCTCTAC 13409

QY 190 ACGTGTGAGTGGAGGCTTCCAGAGCAGCAGCCAGCCCTC 228
Db 13410 ACCTGCCAGTCTGGAGCAGCCCCCGGAGCCCAAGCCTC 13448

RESULT 14
AAS28908/c
ID AAS28908 standard; DNA; 20565 BP.
XX
AC AAS28908;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immunoglobulin encoding genomic DNA SEQ ID No 270.
XX
XX Immunoglobulin; signal transduction pathway protein; cancer; ds;
XX antisense therapy; gene therapy; neurological disorder; renal disorder;
XX cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
XX reproductive disorder; immune system disorder; proliferative disorder;
XX muscular disorder.
XX
```

OS Homo sapiens.
XX WO200155315-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001336.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226275P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229533P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236357P.
PR 29-SEP-2000; 2000US-0236388P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA

PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240360P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0256179P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.

Disclosure; SEQ ID NO 6045; 1297pp + Sequence Listing; English.

XX

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention

SQ Sequence 20565 BP; 4236 A; 6155 C; 5760 G; 4414 T; 0 U; 0 Other;

Query Match

15.1%; Score 60.6; DB 4; Length 20565;

Best Local Similarity 54.8%; Pred. No. 0.00013;

Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 TGGGTGGCGCCAGGGGAGGACGTGGAGCTGGCTGTGAGCTGTCCAGGGGGGAAACGCC 69

DB 13230 TGTGTGGCCCCCTGGGGAGCCAGTGTGCTGAGCTGTGAACCTGCCGGCTGGCGCCCC 13289

QY 70 GTGCACCTGGCTGAAAGGACAGGAAGGCCATCCGCAAGAGCCAGAAAGTATGATGTGCTGTC 129

DB 13290 GTGGCTTGGAGCCACAATGGAGGCCCGTGCAGGAGGGCGAGGSCCTAGAGCTCCATGCC 13349

QY 130 GAGGGCAGGATGGCCATGCTGGTTCATCGCGGGGCCCTGCTCAAGGAGCGGGGCGAGTAC 189

DB 13350 GAGGGCCCCCGCCAGTCTCTGCATCCAGGCTGCAGGCCCCAGCCATGCAGGGCTCTAC 13409

QY 190 ACGTGTGAGGTGGAGGCTTCCAGAGCACAGCCAGCCCTC 228

DB 13410 ACCTGCCAGTCTGGAGCAGCCCCCGAGCCCCCAAGCCTC 13448

Search completed: March 20, 2005, 18:55:02

Job time : 234.766 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 20, 2005, 13:24:10 ; Search time 13537.8 Seconds
(without alignments)
10741.310 Million cell updates/sec

Title: US-10-077-130-6_COPY_19500_22500
Perfect score: 3001
Sequence: 1 cctacggagcagaactggg.....ccactggagcaagacgga 3001

Scoring table:
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba.*
2: gb_hvg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3001	100.0	4936	6	AX207401 Sequence
2	2999.4	99.9	5007	6	AX039410 Sequence
3	2999.4	99.9	7928	6	AX039412 Sequence
4	2999.8	99.9	5207	6	AR255867 Sequence
5	2997.8	99.9	5207	6	AR442826 Sequence
6	2997.8	99.9	5207	6	AR442827 Sequence
7	2997.8	99.9	5207	6	AX430856 Sequence
8	2997.8	99.9	5207	6	AX430858 Sequence
9	2996.2	99.8	5454	6	AX504255 Sequence
10	2643.4	88.1	4175	6	AX039409 Sequence
11	2561	85.3	4041	9	AB046859 Homo sapi
12	2070.4	69.0	62164	9	AL670729 Human DNA
13	1738.8	57.9	18524	6	CQ730656 Sequence
14	1706.8	56.9	3225	6	AX642869 Sequence
c 15	1517	50.5	174612	2	AC023889 Homo sapi
16	1512.4	50.4	4740	10	AY603754 Mus muscu
17	1175.2	39.2	1697	6	CQ731022 Sequence
c 18	1082.8	35.1	164766	2	AC026657 Homo sapi
c 19	835	27.8	103129	10	AL662809 Mouse DNA

c 20	835	27.8	258622	2	AC020877	Mus muscu
c 21	830.6	27.7	260998	2	AC099089	Rattus no
22	643	21.4	250087	2	AC096931	Rattus no
23	615.4	20.5	164766	2	AC026657	Homo sapi
24	325.8	10.9	174612	2	AC023889	Homo sapi
25	235.2	7.8	5923	5	AY603753	Danio rer
26	233.2	7.4	6726	9	AB037718	Homo sapi
27	223.2	7.4	7710	6	AX056397	Sequence
28	223.2	7.4	7789	6	AX056397	Sequence
29	223.2	7.4	9282	9	AY603755	Homo sapi
30	203.4	6.8	10801	10	AF215896	Sequence
31	196.2	6.5	6847	6	CQ714785	Sequence
32	175	5.8	1449	5	BX934738	Gallus ga
33	172	5.7	1410	10	AB007143	Mus muscu
34	172	5.7	1429	6	AR076190	Sequence
35	172	5.7	1429	6	AR124103	Sequence
36	172	5.7	1429	6	E23385	DNA encodin
37	170.2	5.7	1365	12	AY335738	Synthetic
38	170.2	5.7	2055	6	CQ873886	Sequence
39	170.2	5.7	2055	9	AB022341	Homo sapi
40	170.2	5.7	2079	6	CQ782734	Sequence
41	170.2	5.7	2079	6	BD127287	Primer fo
42	170.2	5.7	2079	9	AK074799	Homo sapi
43	170.2	5.7	2118	9	AK097643	Homo sapi
44	168.6	5.6	2105	6	BD176607	Method of
45	168.6	5.6	2105	6	CQ873976	Sequence

ALIGNMENTS

RESULT 1
AX207401
LOCUS AX207401 4936 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 14 from Patent WO0155356.
ACCESSION AX207401
VERSION AX207401.1 GI:15395213
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Plowman, G., Whyte, D., Manning, G., Sudarsanam, S. and Martinez, R.
AUTHORS Human protein kinases and protein kinase-like enzymes
TITLE Patent: WO 0155356-A 14 02-AUG-2001;
JOURNAL Sugen, Inc. (US)
FEATURES
Location/Qualifiers
source
1..4936
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN	Query Match	Best Local Similarity	Matches 3001;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	CCTACGAGCAGAACTCGGGCCAGGCATACAGGAGCGAGACATCTCTGGCCGCTGAG	60								
Db	450	CCTACGAGCAGAACTCGGGCCAGGCATACAGGAGCGAGACATCTCTGGCCGCTGAG	509								
Qy	61	CCACCCGCTGTCACGGGGCTGTGGACACAGTTTGAGACCCGCAAGACCTCATCTCAT	120								
Db	510	CCACCCGCTGTCACGGGGCTGTGGACACAGTTTGAGACCCGCAAGACCTCATCTCAT	569								
Qy	121	CCTGAGCTGTGCTCATCCGAGGCTCTGGACCGCTGTACAGAGAGGGCGTGTGAC	180								
Db	570	CCTGAGCTGTGCTCATCCGAGGCTCTGGACCGCTGTACAGAGAGGGCGTGTGAC	629								
Qy	181	GGAGCCGAGGTCAAGGTCTCATCCAGAGCTGTGTGGAGGGGTGCTACTTCTGCACAG	240								
Db	630	GGAGCCGAGGTCAAGGTCTCATCCAGAGCTGTGTGGAGGGGTGCTACTTCTGCACAG	689								

QY 241 CCATGGCGTTCTCCACCTGACATAAAGCCCTCTAACAATCTGATGTCATCCTGCCCCG 300
Db 690 CCATGGCGTTCTCCACCTGACATAAAGCCCTCTAACAATCTGATGTCATCCTGCCCCG 749
QY 301 GGAAGACATTAAATATCTGGCACTTTGGCTTTGGCCAGAAATCACCCAGAGAGTGCA 360
Db 750 GGAAGAATTAAATATCTGGCACTTTGGCTTTGGCCAGAAATCACCCAGAGAGTGCA 809
QY 361 GTTCAGCCAGTACGGCTCCCTGAGTTGCTCTCCCGGAGATCATCAGCAGAACCTGT 420
Db 810 GTTCAGCCAGTACGGCTCCCTGAGTTGCTCTCCCGGAGATCATCAGCAGAACCTGT 869
QY 421 GAGGGAAGCCTCCGAATTTGGCCATGGGTGTCACTCTCTACCTCAGCCTGACCTGTCTC 480
Db 870 GAGCGAAGCCTCCGAATTTGGCCATGGGTGTCACTCTCTACCTCAGCCTGACCTGTCTC 929
QY 481 ATCCCCATTTGGCGGAGAGTACCGTGGCCACCTCTGAGCCACCTCTCTGAAAGCGGT 540
Db 930 ATCCCCATTTGGCGGAGAGTACCGTGGCCACCTCTCTGAAAGCGGTCTGGAGGGCGGT 989
QY 541 GTCATGGAGCAGCCCATGGCTGCCACCTCAGCGAAGACGCCAAAGACTTTCATCAAGGC 600
Db 990 GTCATGGAGCAGCCCATGGCTGCCACCTCAGCGAAGACGCCAAAGACTTTCATCAAGGC 1049
QY 601 TAGCCTGCAGAGAGCCCTCAGGCCGGCTTAGTGGGCCAGTGCCTCTCCACCCCTG 660
Db 1050 TAGCCTGCAGAGAGCCCTCAGGCCGGCTTAGTGGGCCAGTGCCTCTCCACCCCTG 1109
QY 661 GTTCTCTGAAATCATGCTCGGAGAGGCCACCTTCAATCAACCAAGAGCTCAAGTT 1169
Db 1110 GTTCTCTGAAATCATGCTCGGAGAGGCCACCTTCAATCAACCAAGAGCTCAAGTT 1169
QY 721 CTTCTGGCCGGAAGTCGTGCGAGCTTCTCTGATGAGCTACAAAGTCCATCTGGTGAT 780
Db 1170 CTTCTGGCCGGAAGTCGTGCGAGCTTCTCTGATGAGCTACAAAGTCCATCTGGTGAT 1229
QY 781 GCGCTCCATCTCCTGAGCTGCTGGGGGCCACCGACAGCCCTCTCTGGCGGTAGCCCG 840
Db 1230 GCGCTCCATCTCCTGAGCTGCTGGGGGCCACCGACAGCCCTCTCTGGCGGTAGCCCG 1289
QY 841 GCACCTCTCAGGGACACTGTGTGGCTCTCCAGTTTCTCTCTCTCTCTGACAAAGAGCT 900
Db 1290 GCACCTCTCAGGGACACTGTGTGGCTCTCTCAAGTTCTCTCTCTCTGACAAAGAGCT 1349
QY 901 CGCCCCATTTGCCGGCTTAAGTCACTGCAACCTTCCCGGTGACACTCACCACTGCT 960
Db 1350 CGCCCCATTTGCCGGCTTAAGTCACTGCAACCTTCCCGGTGACACTCACCACTGCT 1409
QY 961 GCACCCCGGGCTTCTCTGCGGCCCTCGGCCAGCTGCTCCGAGGGTCCGCGGCCACCGGCCG 1080
Db 1410 GCACCCCGGGCTTCTCTGCGGCCCTCGGCCAGCTGCTCCGAGGGTCCGCGGCCACCGGCCG 1469
QY 1021 GCGCTCCACCGAGGCCCTCGCTCGCTGCTATCTCCGAGGGTCCGCGGCCACCGGCCG 1080
Db 1470 GCGCTCCACCGAGGCCCTCGCTCGCTGCTATCTCCGAGGGTCCGCGGCCACCGGCCG 1529
QY 1081 CCAGGGCTGCTGCTCCCGGCAAGAGTCACTCGGAGCTGTTTACCAACGAGCGGTGA 1140
Db 1530 CCAGGGCTGCTGCTCCCGGCAAGAGTCACTCGGAGCTGTTTACCAACGAGCGGTGA 1589
QY 1141 GAGCCTCTGAGCA CGGGGCCCTTGGCCCGGGAGCAGCGCACCCGGCCCGGGCGGGCA 1200
Db 1590 GAGCCTCTGAGCA CGGGGCCCTTGGCCCGGGAGCAGCGCACCCGGCCCGGGCGGGCA 1649
QY 1201 CCTGCTGAAGGGCGGTATCATTTGGGGGGCTGCCAGGCTTGGCGAGCACTGATGGA 1260
Db 1650 CCTGCTGAAGGGCGGTATCATTTGGGGGGCTGCCAGGCTTGGCGAGCACTGATGGA 1709
QY 1261 GCACCGGTGCTGAGGAGGAGGCCCGCAGGAGGAGCAGGCCACCTCTCTGGCCAAAGC 1320
Db 1710 GCACCGGTGCTGAGGAGGAGGCCCGCAGGAGGAGCAGGCCACCTCTCTGGCCAAAGC 1769

QY 1321 CCCCTCATTTGAGAGCTGCCCTCCGGCTGCTGCTCTGGCACCCACACTTTGGGCCCTTGGCCA 1380
Db 1770 CCCCTCATTTGAGAGCTGCCCTCCGGCTGCTGCTCTGGCACCCACACTTTGGGCCCTTGGCCA 1829
QY 1381 CAGCCACTCCCTTGGAAATGACTCTCCGAGACACCCCGCCCTCTCTCGAGGGCTTGGG 1440
Db 1830 CAGCCACTCCCTTGGAAATGACTCTCCGAGACACCCCGCCCTCTCTCGAGGGCTTGGG 1889
QY 1441 TGAGGCACAGGAGCTGCTTTCAGCCCTCTCGGGGGGCCCTTATCAGGGACATGGGGCA 1500
Db 1890 TGAGGCACAGGAGCTGCTTTCAGCCCTCTCGGGGGGCCCTTATCAGGGACATGGGGCA 1949
QY 1501 CCCTCAGGGCTCCAAAGCAGCTTCCATCACTGCTGGCCACCCAGGCACTGCTCAGCCAGA 1560
Db 1950 CCCTCAGGGCTCCAAAGCAGCTTCCATCACTGCTGGCCACCCAGGCACTGCTCAGCCAGA 2009
QY 1561 GAGGCCATCCCGGACAGCCCTTGGGGGCAGCCAGCCCTTTCTGCCACCCCAAGCAGGG 1620
Db 2010 GAGGCCATCCCGGACAGCCCTTGGGGGCAGCCAGCCCTTTCTGCCACCCCAAGCAGGG 2069
QY 1621 TTCTGCCCCCAGAGGGCTGACGCCCCCAAGAGCTTGGCCCTTCTGCTCAGCCAGGG 1680
Db 2070 TTCTGCCCCCAGAGGGCTGACGCCCCCAAGAGCTTGGCCCTTCTGCTCAGCCAGGG 2129
QY 1681 CTTCCCTCCAGGATCTTGCAAAGAGGCCCTTAGTACCCTCAAGCCCTTCTTGGGACA 1740
Db 2130 CTTCCCTCCAGGATCTTGCAAAGAGGCCCTTAGTACCCTCAAGCCCTTCTTGGGACA 2189
QY 1741 GCCCCAGGCAACCCCTGCCCCCTGCAAGCAAGCCCCCAATTGGACTCTAAGATGGGGCC 1800
Db 2190 GCCCCAGGCAACCCCTGCCCCCTGCAAGCAAGCCCCCAATTGGACTCTAAGATGGGGCC 2249
QY 1801 TGAGAGACTCTCTTCTCTGGAGGCCAAACCCCGCCCTGAGCTTCCCAAGGGTCAGC 1860
Db 2250 TGAGAGACTCTCTTCTCTGGAGGCCAAACCCCGCCCTGAGCTTCCCAAGGGTCAGC 2309
QY 1861 CTCCAGGGAGCTCTTCCCAAGTGAAGCTCCCTCAGGGTGGGCTCTCCAGGTGGGCAC 1920
Db 2310 CTCCAGGGAGCTCTTCCCAAGTGAAGCTCCCTCAGGGTGGGCTCTCCAGGTGGGCAC 2369
QY 1921 AGAGCCTGGCCCTCTCTGATGCGAGGGCTGGACCCAGAGGGCTGAGGATCTGTCCGA 1980
Db 2370 AGAGCCTGGCCCTCTCTGATGCGAGGGCTGGACCCAGAGGGCTGAGGATCTGTCCGA 2429
QY 1981 CTCCACACCCACTTTCAGCGGCTCAGGAAACAGGGGACCATGCGCAAGTTCTCCCTGGG 2040
Db 2430 CTCCACACCCACTTTCAGCGGCTCAGGAAACAGGGGACCATGCGCAAGTTCTCCCTGGG 2489
QY 2041 TGGTCGGGGGGCTACGCAAGGGTGGCTGCTATGTCACCTTTGCTTTGGTGGAGATGC 2100
Db 2490 TGGTCGGGGGGCTACGCAAGGGTGGCTGCTATGTCACCTTTGCTTTGGTGGAGATGC 2549
QY 2101 AGGGGCACTGCTGGGGCAGGGGCCCATGTTGGGGCAGGATAGCTTGGGCTGTGTCCAGTTC 2160
Db 2550 AGGGGCACTGCTGGGGCAGGGGCCCATGTTGGGGCAGGATAGCTTGGGCTGTGTCCAGTTC 2609
QY 2161 GGAGGAGGAGGAGCAGGAGGCCAGGGCTGAGTCCCAAGTCCGAGGAGCAGCAGAGGC 2220
Db 2610 GGAGGAGGAGGAGCAGGAGGCCAGGGCTGAGTCCCAAGTCCGAGGAGCAGCAGAGGC 2669
QY 2221 CAGGGCTGAGAGCCCACTGCCCCCAGGTCAAGGGCTGTGCTGTAGGGTGGCAGGGC 2280
Db 2670 CAGGGCTGAGAGCCCACTGCCCCCAGGTGAGTCCCAAGGGCTGTGCTGTAGGGTGGCAGGGC 2729
QY 2281 TCCCAACAGGAGCTCTCCAGAGCCCACTGAGGAGGACATCGGGCAGGCTCCTCGGT 2340
Db 2730 TCCCAACAGGAGCTCTCCAGAGCCCACTGAGGAGGACATCGGGCAGGCTCCTCGGT 2789
QY 2341 GCAGATCCGGGACCTGTGTCAGGTGATGCGAGGCGGCCGACAAATATCCCTGGACATTTTC 2400
Db 2790 GCAGATCCGGGACCTGTGTCAGGTGATGCGAGGCGGCCGACAAATATCCCTGGACATTTTC 2849
QY 2401 CGAGGTGGACCCCGCTACCTCAACCTCTCAGACTGTACGATATCAAGTACCTCCCATTT 2460

Db 2850 CGAGTGAACCCCGCTACCTCACTCTCAGACCTGTACGATATCAAGTACTCCCAT 2909
Qy 2461 CGAGTTTATGATCTTTCAGAAAGTCCCCAAGTCCCTCAGCAGAGCCGCCCTCCCCCAT 2520
Db 2910 CGAGTTTATGATCTTTCAGAAAGTCCCCAAGTCCCTCAGCAGAGCCGCCCTCCCCCAT 2969
Qy 2521 GCCTGAGGAGGAGCTGCGCGAGTCTCCGAGGCCACGCTGGCCCTGGCCAGGTGAAGTGGG 2580
Db 2970 GCCTGAGGAGGAGCTGCGCGAGTCTCCGAGGCCACGCTGGCCCTGGCCAGGTGAAGTGGG 3029
Qy 2581 CCCCCACGAGGAGCTGGAGATCACAGAGGAGTCAAGAGTGTGGAGCGGCTGCTGGCAGA 2640
Db 3030 CCCCCACGAGGAGCTGGAGATCACAGAGGAGTCAAGAGTGTGGAGCGGCTGCTGGCAGA 3089
Qy 2641 GCCTGCGGTGGCAGGAGCGCAAGTGTCTCGCGCTCACAGCGCTCTTCCACTTCCC 2700
Db 3090 GCCTGCGGTGGCAGGAGCGCAAGTGTCTCGCGCTCACAGCGCTCTTCCACTTCCC 3149
Qy 2701 TGGGAGGACCTGCGCTGGATGAGCGCTGCAGAGCTGGGGCTGCTGAGAGAGTGAAGGC 2760
Db 3150 TGGGAGGACCTGCGCTGGATGAGCGCTGCAGAGCTGGGGCTGCTGAGAGAGTGAAGGC 3209
Qy 2761 CTCCTGGAGACATCTCCCGATCTCGAGGCGCAGCGGAAAGTCTGGAGAGAGGG 2820
Db 3210 CTCCTGGAGACATCTCCCGATCTCGAGGCGCAGCGGAAAGTCTGGAGAGAGGG 3269
Qy 2821 GCCCCCCAGGAGAACCCAGGCGCTTCTTCTTCCGGCTCTCAGGTCTGAAGAGCTGGGA 2880
Db 3270 GCCCCCCAGGAGAACCCAGGCGCTTCTTCTTCCGGCTCTCAGGTCTGAAGAGCTGGGA 3329
Qy 2881 CCGAGCGCGACATCTTAAGGAGCTCTCAGATGAGAGTGTGTCTTGGGCGAGTCACT 2940
Db 3330 CCGAGCGCGACATCTTAAGGAGCTCTCAGATGAGAGTGTGTCTTGGGCGAGTCACT 3389
Qy 2941 GACACTGGCTGCCAGGTGTGAGCCAGCCAGCTGCCAGGCCACTGGAGCAAGAGCG 3000
Db 3390 GACACTGGCTGCCAGGTGTGAGCCAGCCAGCTGCCAGGCCACTGGAGCAAGAGCG 3449
Qy 3001 A 3001
Db 3450 A 3450

RESULT 2
AX039410
LOCUS AX039410 5007 bp DNA linear PAT 18-NOV-2000
DEFINITION Sequence 3 from Patent WO0063381.
ACCESSION AX039410
VERSION AX039410.1 GI:11229478
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
Zeng, W., Stanton, L. and Kong, H.
AUTHORS
TITLE Mammalian protein with putative function in signal transduction
JOURNAL Patent: WO 0063381-A 3 26-OCT-2000;
SCIOS INC. (US)
FEATURES
Location/Qualifiers
1..5007
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
97...4929
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC16625.1"
/db_xref="GI:11229479"
/translation="MQVTIEDVQAOGTGTAQFEATIEGDPOPSVTWYKDSVOLVDSTR
LSQOQEGTYSIVLRHASKDAGVYVTLAQTGGVLCRAELLVLGDNEDPSEKQSH
RKLSHFYEVKEEIGRGVGFVKRVQHGKNGKILCAKFIPLNSRTRAQYRERDILAA

ORIGIN

Query Match 99.9%; Score 2999.4; DB 6; Length 5007;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CCTACGAGCAGAACTCGGGCCAGGCATACAGGAGCGAGACATCTTGGCCGCGCTGAG 60
Db 522 CCTACGAGCAGAACTCGGGCCAGGCATACAGGAGCGAGACATCTTGGCCGCGCTGAG 581
Qy 61 CCACCCGCTGCTCAGCGGGCTGCTGGACCGCTTTGAGACCCGACAGACCTCATCTCAT 120
Db 582 CCACCCGCTGCTCAGCGGGCTGCTGGACCGCTTTGAGACCCGACAGACCTCATCTCAT 641
Qy 121 CTGGAGCTGTGCTCATCCGAGGAGCTGCTGGACCGCTGTACAGAAAGGGCGTGTGAC 180
Db 642 CTGGAGCTGTGCTCATCCGAGGAGCTGCTGGACCGCTGTACAGAAAGGGCGTGTGAC 701
Qy 181 GGAGCCGAGGTCAAGGTCTACATCCAGCAGCTGTGTGAGGGGTGCACTACCTGACACAG 240
Db 702 GGAGCCGAGGTCAAGGTCTACATCCAGCAGCTGTGTGAGGGGTGCACTACCTGACACAG 761
Qy 241 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTTAACATCTCTGATGTTGTCATCTCCCG 300
Db 762 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTTAACATCTCTGATGTTGTCATCTCCCG 821
Qy 301 GGAGACATTAATAATCTCGACCTTTGGCTTTGGCCAGAAACATCAACCCAGACAGAGCTGCA 360
Db 822 GGAGACATTAATAATCTCGACCTTTGGCTTTGGCCAGAAACATCAACCCAGACAGAGCTGCA 881
Qy 361 GTTCAGCAGTACCGCTCCCTGAGTTCGTCTTCCCGCAGATCATCCAGCAGAAACCTGT 420
Db 882 GTTCAGCAGTACCGCTCCCTGAGTTCGTCTTCCCGCAGATCATCCAGCAGAAACCTGT 941
Qy 421 GAGCGAAGCTCCGACATTTGGGCCATTTGGGCCATTTGGGCCATTTGGGCCATTTGGGCC 480
Db 942 GAGCGAAGCTCCGACATTTGGGCCATTTGGGCCATTTGGGCCATTTGGGCCATTTGGGCC 1001
Qy 481 ATCCCCATTTGCGCGCAGAGTGAACCGTGCACACCTCTCTGAACGCTCTTGGAGGGCGCT 540
Db 1002 ATCCCCATTTGCGCGCAGAGTGAACCGTGCACACCTCTCTGAACGCTCTTGGAGGGCGCT 1061
Qy 541 GTTCAGGAGCAGCCCCCATGGCTGCCACCTCAGCGAAGACGCGCAAGACTTTCATCAAGGC 600
Db 1062 GTTCAGGAGCAGCCCCCATGGCTGCCACCTCAGCGAAGACGCGCAAGACTTTCATCAAGGC 1121
Qy 601 TAGCGTCGACGAGAGGCCCTCAGGGCCGCGCTAGTGGCGCCCGCTAGTGGCGCCCGCTCTCCCA 660

Db	4163	CTCTCTGGCCCGAAGTGCCTGGCAGCGTTCTCTGTATGAGCTACAAGTCCATCTCTGGTGAAT	4222
Qy	781	GGCTCTCATCTCTAGACTGCTCGGGGCCACCCGACAGCCCTCTCTCGCGCTAGCCCG	840
Db	4223	GGCTCCATCTCTAGACTGCTGGGGGCCACCCGAGAGCCCTCTCTCTGGCGTAGCCCG	4282
Qy	841	GCACCTCTGAGAGGACACTGGTGGCTCTCTCAGTTCTCTCTCTCTGTGACAAAGACT	900
Db	4283	GCACCTCTGAGAGGACACTGGTGGCTCTCTCAGTTCTCTCTCTCTCTGTGACAAAGACT	4342
Qy	901	CGCCCCATTGTGCCCGGGCTAAAGTCACTGCACCCCTCCCGGTGACACACTCACCACTGCT	960
Db	4343	CGCCCCATTGTGCCCGGGCTAAAGTCACTGCACCCCTCCCGGTGACACACTCACCACTGCT	4402
Qy	961	GCACCCCCGGGGTTCTTGCGCCCTCTCGGCCAGCTGCTGAGGAAGCGAGGCGAGTGA	1020
Db	4403	GCACCCCCGGGGTTCTTGCGCCCTCTCGGCCAGCTGCTGAGGAAGCGAGGCGAGTGA	4462
Qy	1021	GGCTTCCACCGAGGCCCGAGCTCGCGCTGCATCTCCGAGGGTCCGGGGCACCGGCCCG	1080
Db	4463	GGCTTCCACCGAGGCCCGAGCTCGCGCTGCATCTCCGAGGGTCCGGGGCACCGGCCCG	4522
Qy	1081	CCAGGGCTGCGTGCCTCCCGGCAAGCGTCACTCCGAGCGTGTCTTACCAACAGGCGGGTGA	1140
Db	4523	CCAGGGCTGCGTGCCTCCCGGCAAGCGTCACTCCGAGCGTGTCTTACCAACAGGCGGGTGA	4582
Qy	1141	GAGCCCTGAGCACGGGGCCCTGGCCCGGGAGAGCAGCGGGACCCGGCCCGGGCGGCA	1200
Db	4583	GAGCCCTGAGCACGGGGCCCTGGCCCGGGAGAGCAGCGGGACCCGGCCCGGGCGGCA	4642
Qy	1201	CTGTCTGAAGCGGGGTACATTGGCGGGGGCGCTGCCAGGGCTGCGGAGGCACTGATGA	1260
Db	4643	CTGTCTGAAGCGGGGTACATTGGCGGGGGCGCTGCCAGGGCTGCGGAGGCACTGATGA	4702
Qy	1261	GCACCGCTGTGTGAGAGGAGGGCGCCGAGGAGGAGCAGGCCACCTCTGTGCCAAAGC	1320
Db	4703	GCACCGCTGTGTGAGAGGAGGGCGCCGAGGAGGAGCAGGCCACCTCTGTGCCAAAGC	4762
Qy	1321	CCCTCATTTGAGACTGCGCTTCCGGCTGCTGCTCTGGCAGCCCACTTGGCCCTTGCCCA	1380
Db	4763	CCCTCATTTGAGACTGCGCTTCCGGCTGCTGCTCTGGCAGCCCACTTGGCCCTTGCCCA	4822
Qy	1381	CAGCCACTCTTGAAACATGACTCTCGAGACCCCCCGCCCTCTCTCGAGGCGCTGCGG	1440
Db	4823	CAGCCACTCTTGAAACATGACTCTCGAGACCCCCCGCCCTCTCTCGAGGCGCTGCGG	4882
Qy	1441	TGAGGCACAGCACTGCTTTCAGCCCCCTCCGGGGGGGCCCTATCAGGGACATGGGCA	1500
Db	4883	TGAGGCACAGCACTGCTTTCAGCCCCCTCCGGGGGGGCCCTATCAGGGACATGGGCA	4942
Qy	1501	CCCTCAGGGGTCCAAGCAGCTTCATCTCATGTGGTGGCCACCCAGGCACTGCTCAGCCAGA	1560
Db	4943	CCCTCAGGGGTCCAAGCAGCTTCATCTCATGTGGTGGCCACCCAGGCACTGCTCAGCCAGA	5002
Qy	1561	GAGGCCATCCCGAGACAGCCCTTGGGGGAGCAGAGCCCTTCTTGGCACCACAGAGGG	1620
Db	5003	GAGGCCATCCCGAGACAGCCCTTGGGGGAGCAGAGCCCTTCTTGGCACCACAGAGGG	5062
Qy	1621	TTCTGCCCCCAGGAGGGCTGCAAGCCCCACCCAGCAGTTGCCCCATGCCCCCTCTGGCTC	1680
Db	5063	TTCTGCCCCCAGGAGGGCTGCAAGCCCCACCCAGCAGTTGCCCCATGCCCCCTCTGGCTC	5122
Qy	1681	CTTCCCTCCAGGATCTTGAAAGAGGCGCCTTAGTACCTCAAGCCCTCTTTGGGACA	1740
Db	5123	CTTCCCTCCAGGATCTTGAAAGAGGCGCCTTAGTACCTCAAGCCCTCTTTGGGACA	5182
Qy	1741	GCCCCAGGCACCCCTGCCCCCTGCCAAAGACAAAGCCCCCATTTGGACTCTAAGATGGGGCC	1800
Db	5183	GCCCCAGGCACCCCTGCCCCCTGCCAAAGACAAAGCCCCCATTTGGACTCTAAGATGGGGCC	5242
Qy	1801	TGAGACATCTCTCTTCTGGAGGCGCAAAACCCGGGCCCTCTGAGTCTCCAGGGTCAAG	1860
Db	5243	TGAGACATCTCTCTTCTGGAGGCGCAAAACCCGGGCCCTCTGAGTCTCCAGGGTCAAG	5302

QY	1861	CTCCACGGCAGCTCTTCCCAAGTGA	CTCCTCAGGGTGGGCTCCTCCACAGTGGGCAC	1920
Db	5303	CTCCACGGCAGCTCTTCCCAAGTGA	CTCCTCAGGGTGGGCTCCTCCACAGTGGGCAC	5362
QY	1921	AGAGCCTGGCCCTCCCTGGATGCGGAGGGCTG	AGNACCAGAGAGCTGAGGATCTGTCCGA	1980
Db	5363	AGAGCCTGGCCCTCCCTGGATGCGGAGGGCTG	AGNACCAGAGAGCTGAGGATCTGTCCGA	5422
QY	1981	CTCCACACCCACTTTCAGCGGGCTTCAGGAA	CAGCGCACCATGCGCAAGTTCTCCCTGGG	2040
Db	5423	CTCCACACCCACTTTCAGCGGGCTTCAGGAA	CAGCGCACCATGCGCAAGTTCTCCCTGGG	5482
QY	2041	TGCTTCGGGGGCTACGACNAGGCTGGCTATG	TGBCACTTTGCCCTTTGGTGGAGATGC	2100
Db	5483	TGCTTCGGGGGCTACGACNAGGCTGGCTATG	TGBCACTTTGCCCTTTGGTGGAGATGC	5542
QY	2101	AGGGGCGATGCTGGGGCAGGGGCCCATGCTG	GGGCCCAGGATAGCTTGGCTGGTCCAGATC	2160
Db	5543	AGGGGCGATGCTGGGGCAGGGGCCCATGCTG	GGGGCCAGGATAGCTTGGCTGGTCCAGATC	5602
QY	2161	GGAGGAGGAGCAGCAGGAGAGGCCAGGGCTG	AGTCTCCAGTTCGGAGGAGCAGCAGGAGGC	2220
Db	5603	AGAGGAGGAGCAGCAGGAGAGGCCAGGGCTG	AGTCTCCAGTTCGGAGGAGCAGCAGGAGGC	5662
QY	2221	CAGGGCTGAGAGCCACATGCGCCCAAGTCA	GTGCAAGGCCTGTGCTGAGGTCCGCAAGGC	2280
Db	5663	CAGGGCTGAGAGCCACATGCGCCCAAGTCA	GTGCAAGGCCTGTGCTGAGGTCCGCAAGGC	5722
QY	2281	TTCCACACGAGGCTCTTCAGAGCCACCCTCA	AGCTGTAGTTCGGAGGACATCGGGCAGGTCT	2340
Db	5723	TTCCACACGAGGCTCTTCAGAGCCACCCTCA	AGCTGTAGTTCGGAGGACATCGGGCAGGTCT	5782
QY	2341	GCAGATCCGGACCTGTTCAGTGTATGGGAG	GGCGGCGCGACACAAATATCCCTTGGA	2400
Db	5783	GCAGATCCGGACCTGTTCAGTGTATGGGAG	GGCGGCGCGCGACACAAATATCCCTTGGA	5842
QY	2401	CGAGTGGACCCCGCCTACCTCAACCTCTCA	GAACCTGTACGATATCAAGTACTCTCCAT	2460
Db	5843	CGAGTGGACCCCGCCTACCTCAACCTCTCA	GAACCTGTACGATATCAAGTACTCTCCAT	5902
QY	2461	CGAGTTTATGATCTTCAGGAAGTCCCGTA	AGTCGGCTCAGCCAGCGCCCTCCGCCAT	2520
Db	5903	CGAGTTTATGATCTTCAGGAAGTCCCGTA	AGTCGGCTCAGCCAGCGCCCTCCGCCAT	5962
QY	2521	GGCTGAGGAGGAGCTGGCCGAGTTCCCGA	GGCCACAGTCGCGCTGGCCAGGTGA	2580
Db	5963	GGCTGAGGAGGAGCTGGCCGAGTTCCCGA	GGCCACAGTCGCGCTGGCCAGGTGA	6022
QY	2581	CCCCACGAGGCTGGAGATCA	CAGAGGATGAGGATGTGGACGGCTGCTGGCAGA	2640
Db	6023	CCCCACGAGGCTGGAGATCA	CAGAGGATGAGGATGTGGACGGCTGCTGGCAGA	6082
QY	2641	GGCTGCGTGGGCAGGAAGCGCAAGTGTCT	CGCGCTCAGCCAGCTTCCACTTCCC	2700
Db	6083	GGCTGCGTGGGCAGGAAGCGCAAGTGTCT	CGCGCTCAGCCAGCTTCCACTTCCC	6142
QY	2701	TGGAGAGCACTTCGCTGATGAGCCTTCAG	AGTGGGGCTTGCCTGAGAGAGTGAAGGC	2760
Db	6143	TGGAGAGCACTTCGCTGATGAGCCTTCAG	AGTGGGGCTTGCCTGAGAGAGTGAAGGC	6202
QY	2761	CTCCGTGGAGCATCTCCGGATCTTGAAGG	CGAGGCGGAGGTCTGGAGAGGAGGG	2820
Db	6203	CTCCGTGGAGCATCTCCGGATCTTGAAGG	CGAGGCGGAGGTCTGGAGAGGAGGG	6262
QY	2821	GGCCCCAGGAAGACCCAGGCTTGTCTTCT	TCTCCGGCTCTCAGGTCTGAAGAGCTGGGA	2880
Db	6263	GGCCCCAGGAAGACCCAGGCTTGTCTTCT	TCTCCGGCTCTCAGGTCTGAAGAGCTGGGA	6322
QY	2881	CCGAGCGCCGACATTCCTTAAGGAGGCT	CTCAGATGAGACTGTGCTGCGCCAGTCAGT	2940
Db	6323	CCGAGCGCCGACATTCCTTAAGGAGGCT	CTCAGATGAGACTGTGCTGCGCCAGTCAGT	6382

QY 2941 GACACTGCGCTGCCAGGTGTGTCAGCCAGCCAGCTGCCAGGCGCACCTGGAGCAAGACGG 3000
Db 6393 GACACTGCGCTGCCAGGTGTGTCAGCCAGCCAGCTGCCAGGCGCACCTGGAGCAAGACGG 6442

QY 3001 A 3001
Db 6443 A 6443

RESULT 4
LOCUS AR255867
DEFINITION Sequence 1 from patent US 6482624.
ACCESSION AR255867
VERSION AR255867.1 GI:27305065
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 5207)
AUTHORS Wei, M.-H., Ketchum, K.A., Di Francesco, V. and Beasley, B.M.
TITLE Isolated human kinase proteins, nucleic acid molecules encoding human kinase proteins, and uses thereof
JOURNAL Patent: US 6482624-A 1 19-NOV-2002;
FEATURES Location/Qualifiers
source 1..5207
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.9%; Score 2997.8; DB 6; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCTACGAGCAGAACTCGGGCCCGCCAGGCATACAGGAGCGAGACATCTCTGGCCGCGCTGAG 60
Db 699 CCTACGAGCAGAACTCGGGCCCGCCAGGCATACAGGAGCGAGACATCTCTGGCCGCGCTGAG 758

QY 61 CCACCCGCTGTACGGGGCTGTGGACAGATTTGAGACCGCCAGAACCTCATCTCAT 120
Db 759 CCACCCGCTGTGTACGGGGCTGTGGACAGATTTGAGACCGCCAGAACCTCATCTCAT 818

QY 121 CCTGAGCTGTCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAGGGGGTGTGAC 180
Db 819 CCTGAGCTGTCTCATCCGAGGAGCTGTGGACCGCTGTACAGGAGGGGGTGTGAC 878

QY 181 GGAGCCGAGGTCAAGGTCTACATCCAGCAGCTGTGTGGAGGGCTGTACTTCTGCACAG 240
Db 879 GGAGCCGAGGTCAAGGTCTACATCCAGCAGCTGTGTGGAGGGCTGTACTTCTGCACAG 938

QY 241 CCATGGCGTTCTCCACCTGGACATAAAGCCCTTAACATCTCTGATGGTGCACTCTGCCCG 300
Db 939 CCATGGCGTTCTCCACCTGGACATAAAGCCCTTAACATCTCTGATGGTGCACTCTGCCCG 998

QY 301 GGAAGACATTAAGTCTGGACTTTGGCTTTGGCCAGAACATCAACCCAGCAGACTGCA 360
Db 999 GGAAGACATTAAGTCTGGACTTTGGCTTTGGCCAGAACATCAACCCAGCAGACTGCA 1058

QY 361 GTTCAGCAGATACGGCTCCCTCTGAGTTGGTCTCTCCCGAGATCATCCAGCAGAACCTGT 420
Db 1059 GTTCAGCAGATACGGCTCCCTCTGAGTTGGTCTCTCCCGAGATCATCCAGCAGAACCTGT 1118

QY 421 GAGCGAAGCTCCGACATTTGGGCCATGGGTGTCTATCTCTACCTCAGCTGACCTGTCTC 480
Db 1119 GAGCGAAGCTCCGACATTTGGGCCATGGGTGTCTATCTCTACCTCAGCTGACCTGTCTC 1178

QY 481 ATCCCCATTTGGCCGAGAGTGACCTGTGGCCACCTCTCTGAACGCTCTGGAGGGCGGT 540
Db 1179 ATCCCCATTTGGCCGAGAGTGACCTGTGGCCACCTCTCTGAACGCTCTGGAGGGCGGT 1238

QY 541 GTTCATGAGCAGCCCGCATGGCTGCCACCTCAGCGAAGACGCCAAGACTTTCATCAAGC 600
Db 1239 GTTCATGAGCAGCCCGCATGGCTGCCACCTCAGCGAAGACGCCAAGACTTTCATCAAGC 1298

QY 601 TACCTGTCAGAGAGCCCTCTCAGGCGCGGCTAGTGGGCGCCAGTGCCTCTCCACCCCTG 660
Db 1299 TACCTGTCAGAGAGCCCTCTCAGGCGCGGCTAGTGGGCGCCAGTGCCTCTCCACCCCTG 1358

QY 661 GTTCTGAAATCCATGCTGCGGAGGAGGCCACCTTTCATCAACCAAGAGCAGCTCAAGTT 720
Db 1359 GTTCTGAAATCCATGCTGCGGAGGAGGCCACCTTTCATCAACCAAGAGCAGCTCAAGTT 1418

QY 721 CCTCTGCGCCGAAAGTCTGCTGGCAGCGTTCTCTGATGAGCTACAAGTCAATCTCTGGTAT 780
Db 1419 CCTCTGCGCCGAAAGTCTGCTGGCAGCGTTCTCTGATGAGCTACAAGTCAATCTCTGGTAT 1478

QY 781 GCGCTCCATCTCCTGAGCTGCTGCGGGGCCACCGACAGCCCTCCCTCGGGGTAGCCCG 840
Db 1479 GCGCTCCATCTCCTGAGCTGCTGCGGGGCCACCGACAGCCCTCCCTCGGGGTAGCCCG 1538

QY 841 GCACCTCTGACAGGACACTGGTGGCTCTCCAGTTCCTCTCTCTCTCTCTGACACAGAGCT 900
Db 1539 GCACCTCTGACAGGACACTGGTGGCTCTCTCCAGTTCCTCTCTCTCTCTGACACAGAGCT 1598

QY 901 CGCCCATTTGCCCGGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACAATGCT 960
Db 1599 CGCCCATTTGCCCGGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACAATGCT 1658

QY 961 GCACCCCGGGGGCTCTCTGCGGCCCTCGGCCAGCTGCTGAGGAGCGGAGGCCAGTGA 1020
Db 1659 GCACCCCGGGGGCTCTCTGCGGCCCTCGGCCAGCTGCTGAGGAGCGGAGGCCAGTGA 1718

QY 1021 GCGCTCCACAGGCGCCAGCTCGCCCTGCACTCTCCAGGAGTGGCGGCGCACCCGCGCGC 1080
Db 1719 GCGCTCCACAGGCGCCAGCTCGCCCTGCACTCTCCAGGAGTGGCGGCGCACCCGCGCGC 1778

QY 1081 CAGGGCTGCTGCGCCCGGCACAGGCTCATCCGAGCTGTTCTTACACACAGCGGGTGA 1140
Db 1779 CAGGGCTGCTGCGCCCGGCACAGGCTCATCCGAGCTGTTCTTACACACAGCGGGTGA 1838

QY 1141 GAGCCCTGAGCACCGGGGCTTGGCCCGGGGAGGAGGCGGACCCGCGCCCGCGCGGCA 1200
Db 1839 GAGCCCTGAGCACCGGGGCTTGGCCCGGGGAGGAGGCGGACCCGCGCCCGCGCGGCA 1898

QY 1201 CCTGTGAAGGGCGGGTACATTGGCGGGGCGCTGCGAGGCGCTGCGGAGCCACTGATGA 1260
Db 1899 CCTGTGAAGGGCGGGTACATTGGCGGGGCGCTGCGAGGCGCTGCGGAGCCACTGATGA 1958

QY 1261 GCACCGCTGTGGAGGAGGAGCGCGCCAGGAGGAGCAGCCCTCTGCGCCAAAGC 1320
Db 1959 GCACCGCTGTGGAGGAGGAGCGCGCCAGGAGGAGCAGCCCTCTGCGCCAAAGC 2018

QY 1321 CCCCTCATTCGAGACTGCCCTCGGCTGCTGCTGCTGCGACCCCACTTGGGCCCTGGCCA 1380
Db 2019 CCCCTCATTCGAGACTGCCCTCGGCTGCTGCTGCTGCGACCCCACTTGGGCCCTGGCCA 2078

QY 1381 CAGCCACTCTCTGGAACATGACTCTCCAGACACCCCGCGCCCTCTCTCGAGGCGCTGCGG 1440
Db 2079 CAGCCACTCTCTGGAACATGACTCTCCAGACACCCCGCGCCCTCTCTCGAGGCGCTGCGG 2138

QY 1441 TGAGSCACAGCGACTGCTTTCAGCCCTCTCGGGGGGGGGCCCTTATCAGGGACATCGGGCA 1500
Db 2139 TGAGSCACAGCGACTGCTTTCAGCCCTCTCGGGGGGGGGCCCTTATCAGGGACATCGGGCA 2198

QY 1501 CCCTCAGGGCTCCAAAGCAGCTTTCATCCATCTGCTGGGCCACCCAGGACACTGCTCAGCCAGA 1560
Db 2199 CCCTCAGGGCTCCAAAGCAGCTTTCATCCATCTGCTGGGCCACCCAGGACACTGCTCAGCCAGA 2258

QY 1561 GAGGCCATCCCGGAGACGCCCTTGGGGGAGCAGCCCTTCTGCCACCCCAAGCAGGG 1620
Db 2259 GAGGCCATCCCGGAGACGCCCTTGGGGGAGCAGCCCTTCTGCCACCCCAAGCAGGG 2318

QY 1621 TTCTGCCCGCCAGGAGGGCTGCGAGCCCGCCACCCAGCAGTTGCCCCCATGCTCTCTGGCTC 1680
Db 2319 TTCTGCCCGCCAGGAGGGCTGCGAGCCCGCCACCCAGCAGTTGCCCCCATGCTCTCTGGCTC 2378

QY 1681 CTTCCCTCCAGGATCTTGCAAGAGAGGCCCCCTTTAGTACCTCAAGCCCTTCTTGGGACA 1740
DB |||||
QY 2379 CTTCCCTCCAGGATCTTGCAAGAGAGGCCCCCTTTAGTACCTCAAGCCCTTCTTGGGACA 2438
DB |||||
QY 1741 GCCCCAGGACCCCTGCGCCCTGCCAAGCAGAGCCCCCAATTGGAGCTCTAAGATGGGGCC 1800
DB |||||
QY 2439 GCCCCAGGACCCCTGCGCCCTGCCAAGCAGAGCCCCCAATTGGAGCTCTAAGATGGGGCC 2498
DB |||||
QY 1801 TGGAGACATCTCTCTCTCTGGGAGGCCAAAGCCCGGCCCTTGACGTCTCCCAAGGCTCAGC 1860
DB |||||
QY 2499 TGGAGACATCTCTCTCTGGGAGGCCAAAGCCCGGCCCTTGACGTCTCCCAAGGCTCAGC 2558
DB |||||
QY 1861 CTCACAGGAGCTCTTCCAAAGTGAAGTCTCCCTCAGGGTGGGCTCTCTCCAGAGTGGGAC 1920
DB |||||
QY 2559 CTCACAGGAGCTCTTCCAAAGTGAAGTCTCCCTCAGGGTGGGCTCTCTCCAGAGTGGGAC 2618
DB |||||
QY 1921 AGAGCTCGCCCTCTCCCTGGAATGCCGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 1980
DB |||||
QY 2619 AGAGCTCGCCCTCTCCCTGGAATGCCGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 2678
DB |||||
QY 1981 CTCCACACCCACTTTCAGCGGCTCAGGAAACAGGCGACCATGCGCAAGTTCTCCCTGGG 2040
DB |||||
QY 2679 CTCCACACCCACTTTCAGCGGCTCAGGAAACAGGCGACCATGCGCAAGTTCTCCCTGGG 2738
DB |||||
QY 2041 TGGTCGCGGGGCTACGACAGGCTGGCTGGCTATGGCACTTTCCTTGGCTTGGTGGAGATGC 2100
DB |||||
QY 2739 TGGTCGCGGGGCTACGACAGGCTGGCTGGCTATGGCACTTTCCTTGGCTTGGTGGAGATGC 2798
DB |||||
QY 2101 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCCTGGGGCTGTCCCAATC 2160
DB |||||
QY 2799 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCCTGGGGCTGTCCCAATC 2858
DB |||||
QY 2161 GGAGGAGGAGGACGAGGAGGGCCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2220
DB |||||
QY 2859 GGAGGAGGAGGACGAGGAGGGCCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2918
DB |||||
QY 2221 CAGGCTCAGAGCCACCTGCCCCAGGTCAAGTCAAGGCTGTGCTGAGGTCGGAGGGC 2280
DB |||||
QY 2919 CAGGCTCAGAGCCACCTGCCCCAGGTCAAGTCAAGGCTGTGCTGAGGTCGGAGGGC 2978
DB |||||
QY 2281 TCCACACAGGAGCTCTCCAGAGGCCACCCATGGGAGACATCCGGCAGGCTCCCTGGT 2340
DB |||||
QY 2979 TCCACACAGGAGCTCTCCAGAGGCCACCCATGGGAGACATCCGGCAGGCTCCCTGGT 3038
DB |||||
QY 2341 GCAGATCCGGACCTGTCAAGTGTATGGGAGGGCCGACAAATATCCCTGGACATTTTC 2400
DB |||||
QY 3039 GCAGATCCGGACCTGTCAAGTGTATGGGAGGGCCGACAAATATCCCTGGACATTTTC 3098
DB |||||
QY 2401 CGAGGTGACCCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
DB |||||
QY 3099 CGAGGTGACCCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 3158
DB |||||
QY 2461 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCAGAGCCGCGCTCCCCAT 2520
DB |||||
QY 3159 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCAGAGCCGCGCTCCCCAT 3218
DB |||||
QY 2521 GGCTGAGGAGAGTGGCCGAGTTTCCGGAGCCCAAGTGGCCCTGGCCAGGTGAACCTGGG 2580
DB |||||
QY 3219 GGCTGAGGAGAGTGGCCGAGTTTCCGGAGCCCAAGTGGCCCTGGCCAGGTGAACCTGGG 3278
DB |||||
QY 2581 CCCCACAGGCTGGAGATCAGAGGAGTCAAGAGATGAGAGGCTGGAGCGGCTGGGAGAGA 2640
DB |||||
QY 3279 CCCCACAGGCTGGAGATCAGAGGAGTCAAGAGATGAGAGGATGAGAGCGGCTGGGAGAGA 3338
DB |||||
QY 2641 GGCTGCGCTGGGAGGAGCGCAAGTGGTCTCGCGCTCAGCAGAGCTCTTCCACTTCCC 2700
DB |||||
QY 3339 GGCTGCGCTGGGAGGAGCGCAAGTGGTCTCGCGCTCAGCAGAGCTCTTCCACTTCCC 3398
DB |||||
QY 2701 TGGAGGACCTGCGCTGGATGAGCTTCAGAGCTGGGGCTGGCTGAGAGAGTGAAGGC 2760
DB |||||
QY 3399 TGGAGGACCTGCGCTGGATGAGCTTCAGAGCTGGGGCTGGCTGAGAGAGTGAAGGC 3458
DB |||||
QY 2761 CTCCGTGGAGACATCTCCCGATCTTGAAGGGCAGGCCGGAAGGTCTGGAGAGGAGGG 2820
DB |||||

DB 3459 CTCCGTGGAGCACATCTCCCGATCTCAAGGGCAGGCCGGAAGGTCTGGAGAAAGAGGG 3518
QY 2821 GCCCCAGGAGAAAGCCAGGCTTGTCTCCCTCCGGCTCTCAGGTCTTGAAGAGCTGGGA 2880
DB 3519 GCCCCAGGAGAAAGCCAGGCTTGTCTCCCTCCGGCTCTCAGGTCTTGAAGAGCTGGGA 3578
QY 2881 CCAGAGCCGACATCTCTTAAGGGAGCTCTCAGATGAGACTGTGTCTCTGGGCCAGTCAGT 2940
DB 3579 CCAGAGCCGACATCTCTTAAGGGAGCTCTCAGATGAGACTGTGTCTCTGGGCCAGTCAGT 3638
QY 2941 GACACTGGCTTCAGAGTGTCCAGCCAGCAGCTTGCAGGCGACCTTGGAGCAAGAGCG 3000
DB 3639 GACACTGGCTTCAGAGTGTCCAGCCAGCAGCTTGCAGGCGACCTTGGAGCAAGAGCG 3698
QY 3001 A 3001
DB 3699 A 3699

RESULT 5
AR442826
LOCUS AR442826 5207 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 1 from patent US 6670164.
ACCESSION AR442826
VERSION AR442826.1 GI:42670270
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5207)
AUTHORS Wei, M.-H., Ketchum, K.A., Di Francesco, V. and Beasley, E.M.
TITLE Isolated human kinase proteins
JOURNAL Patent: US 6670164-A 1 30-DEC-2003;
FEATURES Location/Qualifiers
source 1..5207
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.9%; Score 2997.8; DB 6; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CCTACGAGCAGAACTCGGGCCAGGCATACAGGGAGCAGACATCTTGGCCGCGCTGAG 60
DB 699 CCTACGAGCAGAACTCGGGCCAGGCATACAGGGAGCAGACATCTTGGCCGCGCTGAG 758
QY 61 CCACCCGCTGGTCAAGGGGCTGTGGACCAAGTTTGAGACCCGCAAGACCCCTCATCTCAT 120
DB 759 CCACCCGCTGGTCAAGGGGCTGTGGACCAAGTTTGAGACCCGCAAGACCCCTCATCTCAT 818
QY 121 CTTGGAGCTGTGTCTATCCGAGGAGCTGTGGACCGGCTGTACAGGAAGGGCGGTGGAC 180
DB 819 CTTGGAGCTGTGTCTATCCGAGGAGCTGTGGACCGGCTGTACAGGAAGGGCGGTGGAC 878
QY 181 GGAGGCGCAGGTCAAGGTCTTACATCCAGCAGCTGGTGGAGGGGCTGCACTACCTGCACAG 240
DB 879 GGAGGCGCAGGTCAAGGTCTTACATCCAGCAGCTGGTGGAGGGGCTGCACTACCTGCACAG 938
QY 241 CCATGGGCTTCTCCACTGGACATAAAGCCCTCTAAACATCTGTATGGTGCATCTCGCCCG 300
DB 939 CCATGGGCTTCTCCACTGGACATAAAGCCCTCTAAACATCTGTATGGTGCATCTCGCCCG 998
QY 301 GGAAGACATTAATAATCTGGACCTTTGGCTTTGGCCAGAAACATCACCCACAGAGCTGCA 360
DB 999 GGAAGACATTAATAATCTGGACCTTTGGCTTTGGCCAGAAACATCACCCACAGAGCTGCA 1058
QY 361 GTTCAGCAGTACGGCTCCCTTGTGAGTTGGTCTCCCGGAGATCATCCAGCAGAAACCTGT 420
DB 1059 GTTCAGCAGTACGGCTCCCTTGTGAGTTGGTCTCCCGGAGATCATCCAGCAGAAACCTGT 1118
QY 421 GAGCGAAGCCCTCCGACATTTGGGCCCATTTGGGTGTCTCTCTACCTCAGCTGACCTGCTC 480
DB |||||

|||||
Db 1119 GAGCGAAGCCTCCGACATTTGGGCCATTTGGGTGTATCTCTACCTCAGCCTGACCTGCTC 1178
Qy 481 ATCCCCATTTGCGCGGAGAGTGACCGTGCACCCCTCTCTGAAACGTCCTGGAGGGGCGCGT 540
Db 1179 ATCCCCATTTGCGCGGAGAGTGACCGTGCACCCCTCTCTGAAAGTCTCTGGAGGGGCGCGT 1238
Qy 541 GTCTGAGGACGCCCCCATTTGGTGCCTCCTCAGCGAAGACGCCCAAGACTTCATCAAGGC 600
Db 1239 GTCTGAGGACGCCCCCATTTGGTGCCTCCTCAGCGAAGACGCCCAAGACTTCATCAAGGC 1298
Qy 601 TACGCTGAGAGAGCCCTCAGGCGCGGCTAGTGCAGCCCAAGTGCCTCTCCACCCCTG 660
Db 1299 TACGCTGAGAGAGCCCTCAGGCGCGGCTAGTGCAGCCCAAGTGCCTCTCCACCCCTG 1358
Qy 661 GTTCTGAAATCCATGCTGCGGAGAGGCCCATTTCAATCAACCAAGCAGCTCAAGTT 720
Db 1359 GTTCTGAAATCCATGCTGCGGAGAGGCCCATTTCAATCAACCAAGCAGCTCAAGTT 1418
Qy 721 CCTCTGCGCGAGTGCCTGGCAGCGTTCTCTGATGAGCTACAAGTCCATCTGGTGTAT 780
Db 1419 CCTCTGCGCGAAGTGCCTGGCAGCGTTCTCTGATGAGCTACAAGTCCATCTGGTGTAT 1478
Qy 781 CGCTTCCATCCTTGAGCTGCTGCGGGGCCCAACCGACAGCCCTCCTCGGCGTAGCCG 840
Db 1479 CGCTTCCATCCTTGAGCTGCTGCGGGGCCCAACCGACAGCCCTCCTCGGCGTAGCCG 1538
Qy 841 GCACCTCTGCAGGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTCAACAGAGCT 900
Db 1539 GCACCTCTGCAGGGACACTGGTGGCTCTCTCAAGTTCTCTCTCTCTCTGACACAGAGCT 1598
Qy 901 CGCCCTGATTCGCGGGCTAAGTCACTGCGACCTCTCCCGGTGACACACTCACCACTGCT 960
Db 1599 CGCCCTCATTTGCGGGTAAAGTCACTGCGACCTCTCCCGGTGACACACTCACCACTGCT 1658
Qy 961 GCACCCCGGGCTTCTGCGGCTCTGCGCAGCTGCTGAGGAAGCGGAGGCCAGTGA 1020
Db 1659 GCACCCCGGGCTTCTGCGGCTCTGCGCAGCTGCTGAGGAAGCGGAGGCCAGTGA 1718
Qy 1021 CGCTTCCACCGAGGCCCTGCGCTGCAATCTCCGAGGTGCGGGGCCACCGGCCG 1080
Db 1719 CGCTTCCACCGAGGCCCTGCGCTGCAATCTCCGAGGTGCGGGGCCACCGGCCG 1778
Qy 1081 CCAGGCTGCTGCGCGCGGACAGGCTCATCGCAGCCTGTTTACACACAGGCGGTGA 1140
Db 1779 CCAGGCTGCTGCGCGCGGACAGGCTCATCGCAGCCTGTTTACACACAGGCGGTGA 1838
Qy 1141 GAGCCTTGAGCAGCGGGGCTTGGCCCGGGAGCAGGGCGCACCGGCCCGGCGGCA 1200
Db 1839 GAGCCTTGAGCAGCGGGGCTTGGCCCGGGAGCAGGGCGCACCGGCCCGGCGGCA 1898
Qy 1201 CTTCTGAAGGGCGGGTACATTTGGGGGGCGCTGCGAGGCTTGGCGAGCCACTGATGA 1260
Db 1899 CTTCTGAAGGGCGGGTACATTTGGGGGGCGCTGCGAGGCTTGGCGAGCCACTGATGA 1958
Qy 1261 GCACCGGCTGTGAGGAGAGGCGCGCAGGGAGGAGCAGGCGACCCCTCTTGGCGCAAGC 1320
Db 1959 GCACCGGCTGTGAGGAGAGGCGCGCAGGGAGGAGCAGGCGACCCCTCTTGGCGCAAGC 2018
Qy 1321 CCCCTCATTTGAGACTGCGCTCGGCTGCTGCGCTGCGGACCCACTTGGGCCCTGGGCA 1380
Db 2019 CCCCTCATTTGAGACTGCGCTCGGCTGCTGCGCTGCGGACCCACTTGGGCCCTGGGCA 2078
Qy 1381 CAGCCACTCTCTGGAACATGACTCTCTCGAGCAGCCCCCGCCCTCTCTCGAGGCGCTGCGG 1440
Db 2079 CAGCCACTCTCTGGAACATGACTCTCTCGAGCAGCCCCCGCCCTCTCTCGAGGCGCTGCGG 2138
Qy 1441 TGAGGCAAGGCACTGCTTCAAGCCCTCTCGGGGGGCGCCCTATATCAGGAGCATGGGCA 1500
Db 2139 TGAGGCAAGGCACTGCTTCAAGCCCTCTCGGGGGGCGCCCTATATCAGGAGCATGGGCA 2198
Qy 1501 CCCTCAGGGCTCCAGGAGCTTCCATCCACTGTTGGCCACCCAGGCACTGCTCAGCCAGA 1560
|||||

Db 2199 CCCTCAGGGCTCCAAGCAGCTTCCATCCACTGGTGGCCACCCAGGCACTGCTCAGCCAGA 2258
Qy 1561 GAGGCCATTCCTCCGACAGCCCTTGGGGGCAAGCCAGCCCTTTTCTGCCACCCAAAGCAGGG 1620
Db 2259 GAGGCCATTCCTCCGACAGCCCTTGGGGGCAAGCCAGCCCTTTTCTGCCACCCAAAGCAGGG 2318
Qy 1621 TTCTGCCCCCAGGAGGGCTGACGCCCCACCCAGCAGTTCGCCCATGCTCTCTGGGCTC 1680
Db 2319 TTCTGCCCCCAGGAGGGCTGACGCCCCACCCAGCAGTTCGCCCATGCTCTCTGGGCTC 2378
Qy 1681 CTTCCCTCCAGGATCTTTGCAAGAGGGCCCCCTTAGTACCTCTCAAGCCCCCTTTCTGGGACA 1740
Db 2379 CTTCCCTCCAGGATCTTTGCAAGAGGGCCCCCTTAGTACCTCTCAAGCCCCCTTTCTGGGACA 2438
Qy 1741 GCCCAGGCAAGCCCTGCTGCTGCAAGAGAGGCCCCCTCATTTGAGCTTAAGATGGGGCC 1800
Db 2439 GCCCAGGCAAGCCCTGCTGCTGCAAGAGAGGCCCCCTCATTTGAGCTTAAGATGGGGCC 2498
Qy 1801 TGGAGACATCTCTCTCTGAGGAGGCCAAACCCGCGCCCTGTCAGTTCCTCCAGGCTCAGC 1860
Db 2499 TGGAGACATCTCTCTCTGAGGAGGCCAAACCCGCGCCCTGTCAGTTCCTCCAGGCTCAGC 2558
Qy 1861 CTCCAGGCGAGCTCTTCCCAAGTGAGCTCCCTCAGGGTGGGCTCTCTCCAGGCTGGGCAC 1920
Db 2559 CTCCAGGCGAGCTCTTCCCAAGTGAGCTCCCTCAGGGTGGGCTCTCTCCAGGCTGGGCAC 2618
Qy 1921 AGAGCTTGGGCCCTCTCTGATGCGAGGGCTGACCCAGAGGCTGAGGATCTGTTCGA 1980
Db 2619 AGAGCTTGGGCCCTCTCTGATGCGAGGGCTGAGCCAGAGGCTGAGGATCTGTTCGA 2678
Qy 1981 CTCCACACCCACTTTCAGCGGCTCAGGACAGGCGACCATGCGCAAGTTCTCCCTGGG 2040
Db 2679 CTCCACACCCACTTTCAGCGGCTCAGGAAACAGGTGACCATGCGCAAGTTCTCCCTGGG 2738
Qy 2041 TGGTTCGCGGGGCTTACGACGCGTGGCTTGTGCACTTTTGCCTTTGCTGGAGATGC 2100
Db 2739 TGGTTCGCGGGGCTTACGACGCGTGGCTTGTGCACTTTTGCCTTTGCTGGAGATGC 2798
Qy 2101 AGGGGGATGCTGGGCGAGGGGCCATGTGGGCGAGATAGCCTGGGCTGTGTCCAGTC 2160
Db 2799 AGGGGGATGCTGGGCGAGGGGCCATGTGGGCGAGATAGCCTGGGCTGTGTCCAGTC 2858
Qy 2161 GGAGGAGGAGCAGAGGAGGCGCAGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 2220
Db 2859 GGAGGAGGAGCAGAGGAGGCGCAGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 2918
Qy 2221 CAGGGCTGAGAGGCCACTTGCCTCCAGGTCAGTGAAGGCTGTGCTGAGGTCGGCAGGCG 2280
Db 2919 CAGGGCTGAGAGGCCACTTGCCTCCAGGTCAGTGAAGGCTGTGCTGAGGTCGGCAGGCG 2978
Qy 2281 TCCACAGGAGCTCTCAGAGCCACCCCATGGAGGACATCGGCGAGGCTCTCCCTGGT 2340
Db 2979 TCCACAGGAGCTCTCAGAGCCACCCCATGGAGGACATCGGCGAGGCTCTCCCTGGT 3038
Qy 2341 GCAGATCCGGGACCTGTTCAGGTGATGCGAGGCGCGGCGACACATATCCCTGGACATTTTC 2400
Db 3039 GCAGATCCGGGACCTGTTCAGGTGATGCGAGGCGCGGCGACACATATCCCTGGACATTTTC 3098
Qy 2401 CGAGGTGACCCCGCTTACCTCAACCTCTCAGACTGTACGATATCAAGTACCTCCCAATT 2460
Db 3099 CGAGGTGACCCCGCTTACCTCAACCTCTCAGACTGTACGATATCAAGTACCTCCCAATT 3158
Qy 2461 CGAGTTTATGATTTTCAAGGAAAGTCCCCAAGTCCGCTCAGCAGAGCGCGCTCCCCCAT 2520
Db 3159 CGAGTTTATGATTTTCAAGGAAAGTCCCCAAGTCCGCTCAGCAGAGCGCGCTCCCCCAT 3218
Qy 2521 GGCTGAGGAGGAGCTGCGCGAGTTCCCGGAGCCCACTGTCGCGCTTGGCCCTGAGTGAACCTGGG 2580
Db 3219 GGCTGAGGAGGAGCTGCGCGAGTTCCCGGAGCCCACTGTCGCGCTTGGCCCTGAGTGAACCTGGG 3278
Qy 2581 CCCCCAGCAGGCTGAGGATCACAGAGGAGTCAAGAGAGTGTGACGCGCTGTCTGGCAGA 2640
Db 3279 CCCCCAGCAGGCTGAGGATCACAGAGGAGTCAAGAGAGTGTGACGCGCTGTCTGGCAGA 3338
|||||

Db 759 CACCCCGCTGGTCA CGGGGCTGTGGACCAAGTTTGAGACCCGCAAGACCTCATCCTCAT 818
Qy 121 CTTGGAGCTGTCTCATCCGAGAGCTGTGTGACCGCTGTACAGGAAGGGGTGTGTGAC 180
Db 819 CTTGGAGCTGTCTCATCCGAGAGCTGTGTGACCGCTGTACAGGAAGGGGTGTGTGAC 878
Qy 181 GGAGGCGAGGTCAAGGTCTACATCCAGCAGCTGTGTGGAGGGGTGCATCTACTGACACAG 240
Db 879 GGAGGCGAGGTCAAGGTCTACATCCAGCAGCTGTGTGGAGGGGTGCATCTACTGACACAG 938
Qy 241 CAGTGGGTCTCCACTGTGACATAAAGCCCTCTAACTCTGTATGTGCATCTCTGCCCG 300
Db 939 CCAATGGCTTCTCCACTGTGACATAAAGCCCTCTAACTCTGTATGTGCATCTCTGCCCG 998
Qy 301 GGAAGACATTAATACTGGACTTTGGCTTTGCCAGAAACATCAACCCAGAGAGCTGCA 1059
Db 999 GGAAGACATTAATACTGGACTTTGGCTTTGCCAGAAACATCAACCCAGAGAGCTGCA 1059
Qy 361 GTTACGCCAGTACGGCTCCCTGAGTTCTCTCCCGAGATCATCCAGCAGAAACCTGT 420
Db 1059 GTTACGCCAGTACGGCTCCCTGAGTTCTCTCCCGAGATCATCCAGCAGAAACCTGT 1118
Qy 421 GAGCGAAGCTCCGACATTTGGCCATGGGTGTGATCTCTTACCTCAGCTGACCTGTCTC 480
Db 1119 GAGCGAAGCTCCGACATTTGGCCATGGGTGTGATCTCTTACCTCAGCTGACCTGTCTC 1178
Qy 481 ATCCCATTTGCGGCGAGGTGACCGTGCACACCTCTCTGAACGCTCTTGGAGGGGCGGT 540
Db 1179 ATCCCATTTGCGGCGAGGTGACCGTGCACACCTCTCTGAACGCTCTTGGAGGGGCGGT 1238
Qy 541 GTCTGAGCAGCCCGATGGCTGCCACCTCAGCGAAGACGCGCAAGACTTTCATCAAGGC 600
Db 1239 GTCTGAGCAGCCCGATGGCTGCCACCTCAGCGAAGACGCGCAAGACTTTCATCAAGGC 1298
Qy 601 TACGCTGACAGAGCCCTCAGCGCCGCGCTAGTGCAGCCGCTCTCCGACCCCTG 660
Db 1299 TACGCTGACAGAGCCCTCAGCGCCGCGCTAGTGCAGCCGCTCTCCGACCCCTG 1358
Qy 661 GTTCTGAAATCCATGCTGGGAGAGGCCCCACTTCAATCAACCAAGCAGCTCAAGTT 720
Db 1359 GTTCTGAAATCCATGCTGGGAGAGGCCCCACTTCAATCAACCAAGCAGCTCAAGTT 1418
Qy 721 CCTCTGCGCCGAGTCTGTCAGCAGCTTCCCTGATGAGCTTACAGTCCATCTCTGGTGTAT 780
Db 1419 CCTCTGCGCCGAGTCTGTCAGCAGCTTCCCTGATGAGCTTACAGTCCATCTCTGGTGTAT 1478
Qy 781 GCGCTCCATCCTGAGCTGCTGCGGGGCCACCGACAGCCCTCCCTCGGCGTAGCCCG 840
Db 1479 GCGCTCCATCCTGAGCTGCTGCGGGGCCACCGACAGCCCTCCCTCGGCGTAGCCCG 1538
Qy 841 GCACTCTGACAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTGACAAACGAGCT 900
Db 1539 GCACTCTGACAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTGACAAACGAGCT 1598
Qy 901 CGCCCCATTTGCCCGGGTAACTCACTGCCACCTCCCGGTGACACACTCAACATGCT 960
Db 1599 CGCCCCATTTGCCCGGGTAACTCACTGCCACCTCCCGGTGACACACTCAACATGCT 1658
Qy 961 GCACCCCGGGGCTTCTTGGCGCCCTCGGCGAGCTGCTGAGGAAGCCGAGGCCAGTGA 1020
Db 1659 GCACCCCGGGGCTTCTTGGCGCCCTCGGCGAGCTGCTGAGGAAGCCGAGGCCAGTGA 1718
Qy 1021 GCGCTTCAACGAGGCCAGCTCGCTGTGANTCTCCGAGGGTGCAGGGCCACCGGCGCG 1080
Db 1719 GCGCTTCAACGAGGCCAGCTCGCTGTGANTCTCCGAGGGTGCAGGGCCACCGGCGCG 1778
Qy 1081 CCAGGGCTGTGTCGCCCGACAGCTCATCCGAGCCTGTCTTACCAACAGCGGGTGA 1140
Db 1779 CCAGGGCTGTGTCGCCCGACAGCTCATCCGAGCCTGTCTTACCAACAGCGGGTGA 1838
Qy 1141 GAGCCCTGAGCA CGGGGCTTGGCCCGGGGAGCAGCGCGCAACCGCCCGCGCGCGCA 1200

Db 1839 GAGCCCTGAGCACGGGGCCCTTGGCCCCGGGGAGCAGGGGGCACCGGGCCCGCGCGGCA 1898
Qy 1201 CTTGCTGAAGGGCGGGTACATTTGCGGGGGCGCTGCCAGGCTGTGCGAGCCACTGATGGA 1260
Db 1899 CTTGCTGAAGGGCGGGTACATTTGCGGGGGCGCTGCCAGGCTGTGCGAGCCACTGATGGA 1958
Qy 1261 GCACCGCTGCTGAGAGGAGGGCGCGCAGGAGGAGCAGGCCACCTCTCTGGGCCAAAGC 1320
Db 1959 GCACCGCTGCTGAGAGGAGGGCGCGCAGGAGGAGCAGGCCACCTCTCTGGGCCAAAGC 2018
Qy 1321 CCCCTCATTTGAGACTGCCCCCTCCGGCTGCTCTCTGGGACCCACCTTGGCCCCCTGGCCA 1380
Db 2019 CCCCTCATTTGAGACTGCCCCCTCCGGCTGCTCTCTGGGACCCACCTTGGCCCCCTGGCCA 2078
Qy 1381 CAGCCACTCCCTTGAAACATGACTCTCCGAGACCCCCCGGCCCTCTCTCGAGGCGCTCGG 1440
Db 2079 CAGCCACTCCCTTGAAACATGACTCTCCGAGACCCCCCGGCCCTCTCTCGAGGCGCTCGG 2138
Qy 1441 TGAGGCACAGCAGCTGCTTTCAGCCCCCTCCGGGGGGCCCCCTATCAGGGACATGGGGCA 1500
Db 2139 TGAGGCACAGCAGCTGCTTTCAGCCCCCTCCGGGGGGCCCCCTATCAGGGACATGGGGCA 2198
Qy 1501 CCCTCAGGGTCCAAAGCAGCTTCCATCCACTGCTGGGCCACCCAGGCACTCTCTCAGCCAGA 1560
Db 2199 CCCTCAGGGTCCAAAGCAGCTTCCATCCACTGCTGGGCCACCCAGGCACTCTCTCAGCCAGA 2258
Qy 1561 GAGGCCATCCCGGACAGCCCTTTGGGGGAGCCAGCCCTTTCTGCCACCCCAAGCAGGG 1620
Db 2259 GAGGCCATCCCGGACAGCCCTTTGGGGGAGCCAGCCCTTTCTGCCACCCCAAGCAGGG 2318
Qy 1621 TTCTGCCCCCAGGAGGGCTGACAGCCCCCACCAGCAGGTTGCCCCCATGCCCTCTCTGGCTC 1680
Db 2319 TTCTGCCCCCAGGAGGGCTGACAGCCCCCACCAGCAGGTTGCCCCCATGCCCTCTCTGGCTC 2378
Qy 1681 CTTTCCCTCAGAGATCTTGCAAGAGGCGCCCTTAGTACCTCAAGCCCTTTCTTGGGACA 1740
Db 2379 CTTTCCCTCAGAGATCTTGCAAGAGGCGCCCTTAGTACCTCAAGCCCTTTCTTGGGACA 2438
Qy 1741 GCCCCAGGCACCCCTGCCCCCTGCCAAAGCAAGCCCCCATTTGGACTCTTAAGATGGGGCC 1800
Db 2439 GCCCCAGGCACCCCTGCCCCCTGCCAAAGCAAGCCCCCATTTGGACTCTTAAGATGGGGCC 2498
Qy 1801 TGAGACATCTCTCTTCTTCTGGAGGCGCAAAACCCGGCCCTCTGAGTTCCCCAGGGTCTCAG 1860
Db 2499 TGAGACATCTCTCTTCTTCTGGAGGCGCAAAACCCGGCCCTCTGAGTTCCCCAGGGTCTCAG 2558
Qy 1861 CTTCCAGGCGAGCTCTTCCCAAGTGAAGTCTCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 1920
Db 2559 CTTCCAGGCGAGCTCTTCCCAAGTGAAGTCTCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 2618
Qy 1921 AGAGCCTGGCCCTCTCTGATGCGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 1980
Db 2619 AGAGCCTGGCCCTCTCTGATGCGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 2678
Qy 1981 CTTCCACACCACTTCTGACGCGCTCAGGAACAGGCGACCATGCGCAAGTTCTCTCCCTGGG 2040
Db 2679 CTTCCACACCACTTCTGACGCGCTCAGGAACAGGCTGACCATGCGCAAGTTCTCTCCCTGGG 2738
Qy 2041 TGGTTCGCGGGGCTACGAGGGCTGGCTGTATGGACCTTTGCTTTGGTGGAGATGC 2100
Db 2739 TGGTTCGCGGGGCTACGAGGGCTGGCTGTATGGACCTTTGCTTTGGTGGAGATGC 2798
Qy 2101 AGGGGCGATGCTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTTGGGCTGTGTCCAGTCT 2160
Db 2799 AGGGGCGATGCTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTTGGGCTGTGTCCAGTCT 2858
Qy 2161 GAGAGAGAGGAGCAGGAGGAGCCAGGGCTGAGTCTCCAGTCTGAGGAGCAGCAGAGGC 2220
Db 2859 GAGAGAGAGGAGCAGGAGGAGCCAGGGCTGAGTCTCCAGTCTGAGGAGCAGCAGAGGC 2918
Qy 2221 CAGGGCTCAGAGCCCACTGCCCCCAGGTCAGTGCAGGGCTGTGTGAGGCTGGCAGGGC 2280
Db 2919 CAGGGCTCAGAGCCCACTGCCCCCAGGTCAGTGCAGGGCTGTGTGAGGCTGGCAGGGC 2978

```
QY 2281 TCCACACGAGGAGCTCTCCAGAGCCCAACCCCATGGAGAGACATCGGGCAGGTCCTCCTGGT 2340
Db 2979 TCCACACGAGGAGCTCTCCAGAGCCCAACCCCATGGAGAGACATCGGGCAGGTCCTCCTGGT 3038
QY 2341 CGAGATCCGGGACCTGTCCAGGTGATCGGAGGCGCCGACACAAATATCCCTGGACATTTTC 2400
Db 3039 CGAGATCCGGGACCTGTCCAGGTGATCGGAGGCGCCGACACAAATATCCCTGGACATTTTC 3098
QY 2401 CGAGTGACACCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db 3099 CGAGTGACACCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 3158
QY 2461 CGAGTTTATGATCTTCAGAAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 2520
Db 3159 CGAGTTTATGATCTTCAGAAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 3218
QY 2521 GCCTGAGGAGGAGCTGGCGGAGTCCCGAGGCGCACGTCGGCCCTGGCCAGGTGAATGGG 2580
Db 3219 GCCTGAGGAGGAGCTGGCGGAGTCCCGAGGCGCACGTCGGCCCTGGCCAGGTGAATGGG 3278
QY 2581 CCCCCACGAGCCTGGAGATCACAGAGGAGTCAAGAGATGTGACGCGCTGTGGCAGA 2640
Db 3279 CCCCCACGAGCCTGGAGATCACAGAGGAGTCAAGAGATGTGACGCGCTGTGGCAGA 3338
QY 2641 GCCTGCCGTGGCAGGAAAGCGCAAGTGTCTCGCCGTCACGAGCCTCTTCCACTTCCC 2700
Db 3339 GCCTGCCGTGGCAGGAAAGCGCAAGTGTCTCGCCGTCACGAGCCTCTTCCACTTCCC 3398
QY 2701 TGGGAGGACACCTGCGCGCTGGATGAGCCTGCAGAGCTGGGGCTGCTGAGAGATGAAGGC 2760
Db 3399 TGGGAGGACACCTGCGCGCTGGATGAGCCTGCAGAGCTGGGGCTGCTGAGAGATGAAGGC 3458
QY 2761 CTCCTGGAGCAGCATCTCCCGGATCCTGAAGGCGAGGCGGAGGTCGTGGAGAGAGGG 2820
Db 3459 CTCCTGGAGCAGCATCTCCCGGATCCTGAAGGCGAGGCGGAGGTCGTGGAGAGAGGG 3518
QY 2821 GCGGCGGAGGAGAGCAGGCTGCTCTCCGCTCTCAGAGCTGGAGAGTGGGA 2880
Db 3519 GCGGCGGAGGAGAGCAGGCTGCTCTCCGCTCTCAGAGCTGGAGAGTGGGA 3578
QY 2881 CCGAGCGCGCATCTCTTAAGGAGCTCTCAGATGAGAGTGTGGTCTGGGCGAGTCAGT 2940
Db 3579 CCGAGCGCGCATCTCTTAAGGAGCTCTCAGATGAGAGTGTGGTCTGGGCGAGTCAGT 3638
QY 2941 GACACTGGCCTGCCAGGTGTGAGCCAGCAGCTGCCAGGCGCACTGGAGCAAGAGCGG 3000
Db 3639 GACACTGGCCTGCCAGGTGTGAGCCAGCAGCTGCCAGGCGCACTGGAGCAAGAGCGG 3698
QY 3001 A 3001
Db 3699 A 3699

RESULT 8
AX430858
LOCUS AX430858 5207 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 3 from Patent WO0240683.
ACCESSION AX430858
VERSION AX430858.1 GI:21655922
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Kechum,K., Beasley,E.M., Wei,M.H. and di Francesco,V.
AUTHORS Isolated human kinase proteins, nucleic acid molecules encoding
TITLE human kinase proteins, and uses thereof
JOURNAL Patent: WO 0240683-A 3 23-MAY-2002;
PE CORP NY (US)
FEATURES
Location/Qualifiers
source 1..5207
```

```
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
```

ORIGIN

```
Query Match 99.9%; Score 2997.8; DB 6; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCTACGGAGCAGAACTCGGGCCCGAGGCATACAGGAGCGAGACATCTCTGGCGCGCTGAG 60
Db 699 CCTACGGAGCAGAACTCGGGCCCGAGGCATACAGGAGCGAGACATCTCTGGCGCGCTGAG 758
QY 61 CACCCCGCTGTCTACGGGGCTGTGGACAGTTTGGACCCGCAAGACCTCTCATCTCAT 120
Db 759 CACCCCGCTGTCTACGGGGCTGTGGACAGTTTGGACCCGCAAGACCTCTCATCTCAT 818
QY 121 CTTGAGAGCTGTCTATCCGAGGAGCTGTCTGACCGCCTGTACAGGAAGGGGTGGTGCAC 180
Db 819 CTTGAGAGCTGTCTATCCGAGGAGCTGTCTGACCGCCTGTACAGGAAGGGGTGGTGCAC 878
QY 181 GGAGCGCAGAGTCAAGGTCTTACATCCAGCAGCTGTGTGGAGGGGTGCACTACCTGCACAG 240
Db 879 GGAGCGCAGAGTCAAGGTCTTACATCCAGCAGCTGTGTGGAGGGGTGCACTACCTGCACAG 938
QY 241 CCATGGCGTTCTCCACCTTGGACATTAAGCCCTCTAAACATCTGTATGTGTGATCTCTGCCG 300
Db 939 CCATGGCGTTCTCCACCTTGGACATTAAGCCCTCTAAACATCTGTATGTGTGATCTCTGCCG 998
QY 301 GGAAGACATTAATAATCTGCGACTTTGGCTTTGGCCAGAAACATCAACCCAGAGAGCTGCA 360
Db 999 GGAAGACATTAATAATCTGCGACTTTGGCTTTGGCCAGAAACATCAACCCAGAGAGCTGCA 1058
QY 361 GTTCAGCGCAGTACCGCTCCCTCGATTCGTCTCTCCCGAGAGATCATCCAGCAGAAACCTGT 420
Db 1059 GTTCAGCGCAGTACCGCTCCCTCGATTCGTCTCTCCCGAGAGATCATCCAGCAGAAACCTGT 1118
QY 421 GAGCGAAGCTTCCGACATTTGGGCGCATGGGTGTCTCTCTACCTCAGCTGACCTGCTC 480
Db 1119 GAGCGAAGCTTCCGACATTTGGGCGCATGGGTGTCTCTCTCTACCTCAGCTGACCTGCTC 1178
QY 481 ATCCCCATTTCCCGCGCAGAGTGACCGTGCACACCTCTCTGAAACGTCCTGGAGGGCGGCT 540
Db 1179 ATCCCCATTTCCCGCGCAGAGTGACCGTGCACACCTCTCTGAAACGTCCTGGAGGGCGGCT 1238
QY 541 GTCATGGAGCAGCCCCATGGCTGCGCCACCTCAGCGAAGACGCCCAAGACTTTCATCAAGGC 600
Db 1239 GTCATGGAGCAGCCCCATGGCTGCGCCACCTCAGCGAAGACGCCCAAGACTTTCATCAAGGC 1298
QY 601 TACGCTGCAGAGAGCCCTCTCAGGCCCGGCTTAGTGTGGGCCCGAGTGTCTCTCCACCCCTG 660
Db 1299 TACGCTGCAGAGAGCCCTCTCAGGCCCGGCTTAGTGTGGGCCCGAGTGTCTCTCCACCCCTG 1358
QY 661 GTTCTCTGAATTCATGCTCGGAGGAGGCCCATCTTCATCAACCAAGAGAGCTCAAGTT 720
Db 1359 GTTCTCTGAATTCATGCTCGGAGGAGGCCCATCTTCATCAACCAAGAGAGCTCAAGTT 1418
QY 721 CCTCTGCGCCGAAAGTCTGCTGCGCAGCGTTCCCTGATGAGCTACAAAGTCCATCTCTGGTGT 780
Db 1419 CCTCTGCGCCGAAAGTCTGCTGCGCAGCGTTCCCTGATGAGCTACAAAGTCCATCTCTGGTGT 1478
QY 781 GCGTCTCATCTCTGAGTGTGTGGGGGCCACCCGACAGCCCTCTCTCTGGCGGTAGCCCG 840
Db 1479 GCGTCTCATCTCTGAGTGTGTGGGGGCCACCCGACAGCCCTCTCTCTGGCGGTAGCCCG 1538
QY 841 GCACCTCTGAGGAGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTCTGACAGAGCT 900
Db 1539 GCACCTCTGAGGAGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTCTGACAGAGCT 1598
QY 901 CGCCCCATTTCCCGCGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACCACCTGCT 960
Db 1599 CGCCCCATTTCCCGCGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACCACCTGCT 1658
```


VERSION	AX504255.1	GI:23386095	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		
AUTHORS	Gururajan, R., Baughn, M.R., Wallia, N.K., Elliott, V.S., Xu, Y., Arvizu, C., Yeo, M.G., Ramkumar, J., Ding, L., Tang, Y.T., Hatalla, A.J., Nguyen, D.B., Gandhi, A.R., Lu, Y., Yue, H., Burford, N., Bandman, O., Tribouley, C.M., Lal, P.G., Recipon, S.A., Lu, D.A., Borowsky, M.L., Thornton, M., Swarnaker, A., Thangavelu, K., Khan, F.A. and Ison, C.H.		
TITLE	Human kinases		
JOURNAL	Patent: WO 0233099-A 44 25-APR-2002;		
FEATURES	Location/Qualifiers		
source	1..5454		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	/note="Incyte ID No: 7638121CB1"		
ORIGIN			
	Query Match 99.8%; Score 2996.2; DB 6; Length 5454;		
	Best Local Similarity 99.9%; Pred No. 0;		
	Matches 2998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
Qy	1	CCTACGGAGCAGAACTCGGGCCCAAGGCATACAGGGAGCGAGACATCCCTGGCGCGCTGAG	60
Db	963	CCTACGGAGCAGAACTCGGGCCCAAGGCATACAGGGAGCGAGACATCCCTGGCGCGCTGAG	1022
Qy	61	CCACCCGCTGTGTCACGGGGCTGTGGACCACTTTGAGACCCCGAAGCCCTCATCTCAT	120
Db	1023	CCACCCGCTGTGTCACGGGGCTGTGGACCACTTTGAGACCCCGAAGCCCTCATCTCAT	1082
Qy	121	CCTGAGCTGTGCTCATCCGAGGAGCTGCTGGACCGCTGTACAGAAAGGGCGTGGTAC	180
Db	1083	CCTGAGCTGTGCTCATCCGAGGAGCTGCTGGACCGCTGTACAGAAAGGGCGTGGTAC	1142
Qy	181	GGAGCCGAGGTCAAGGTCTATCCAGCAGCTGGTGGAGGGGTGCACTACCTGCAAG	240
Db	1143	GGAGCCGAGGTCAAGGTCTATCCAGCAGCTGGTGGAGGGGTGCACTACCTGCAAG	1202
Qy	241	CCATGGCGTTCTCCACCTGGACATAAGCCCTCTAAACATCTGTATGGTGCATCTGCGCG	300
Db	1203	CCATGGCGTTCTCCACCTGGACATAAGCCCTCTAAACATCTGTATGGTGCATCTGCGCG	1262
Qy	301	GGAGACATTAATAATCTCGACTTTGGCTTTGGCCAGAACATCACCCAGCAGAGCTGCA	360
Db	1263	GGAGACATTAATAATCTCGACTTTGGCTTTGGCCAGAACATCACCCAGCAGAGCTGCA	1322
Qy	361	GTTACCCAGTACGGCTCCCTGAGTTGCTCTCCCCGAGATCATCCAGCAGAACCTGT	420
Db	1323	GTTACCCAGTACGGCTCCCTGAGTTGCTCTCCCCGAGATCATCCAGCAGAACCTGT	1382
Qy	421	GAGGAAGCTCCGACATTTGGGCCATGGGTGTATCTCTACCTCAGCCTGACCTGCTC	480
Db	1383	GAGGAAGCTCCGACATTTGGGCCATGGGTGTATCTCTACCTCAGCCTGACCTGCTC	1442
Qy	481	ATCCCCATTTGCGCGAGAGTGACCGTGGCCACCTCTGAACTGCTGGAGGGCGCGT	540
Db	1443	ATCCCCATTTGCGCGAGAGTGACCGTGGCCACCTCTGAACTGCTGGAGGGCGCGT	1502
Qy	541	GTCATGGAGCAGCCCCCATGGTGGCCACTCAGCGAAGACGCGCAAGACTTTCATCAAGGC	600
Db	1503	GTCATGGAGCAGCCCCCATGGTGGCCACTCAGCGAAGACGCGCGCAAGACTTTCATCAAGGC	1562
Qy	601	TACGTCGAGAGAGCCCTCAGGCCCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTG	660
Db	1563	TACGTCGAGAGAGCCCTCAGGCCCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTG	1622
Qy	661	GTTCTGAATTCATGCTCGGAGGAGGCCACCTTCATCAACCAAGCAGCTCAAGTT	720
Db			

Db	1623	GTTCTGAATTCATGCTCGGAGGAGGCCCACTTTCATCAACCAAGCAGCTCAAGTT	1682
Qy	721	CCTCTCGCCCGAAGTGCCTGGCAGCGTTCCCTGATGAGCTACAGTCCATCTCGGTGAT	780
Db	1683	CCTCTCGCCCGAAGTGCCTGGCAGCGTTCCCTGATGAGCTACAGTCCATCTCGGTGAT	1742
Qy	781	GCCTGTCATCCCTGAGCTGCTCGGGGCCCAACCGACAGCCCTCCCTCGGCGTAGCCCG	840
Db	1743	GCCTGTCATCCCTGAGCTGCTCGGGGCCCAACCGACAGCCCTCCCTCGGCGTAGCCCG	1802
Qy	841	GCACCTCTGCAAGGACACACTGGTGGCTCCTCCAGTTCTCTCTCTCTCTGCAAGAGCT	900
Db	1803	GCACCTCTGCAAGGACACACTGGTGGCTCCTCCAGTTCTCTCTCTCTCTGCAAGAGCT	1862
Qy	901	CGCCCCATTTGCCCGGGCTAAGTCACTGCGCACCCTCCCGGTGACACTCACCACCTGCT	960
Db	1863	CGCCCCATTTGCCCGGGCTAAGTCACTGCGCACCCTCCCGGTGACACTCACCACCTGCT	1922
Qy	961	GCACCCCGGGGCTTCTGCGGCCCTCGGGCAGCTGCTCAGGAAGCCGAGGCGCAGTGA	1020
Db	1923	GCACCCCGGGGCTTCTGCGGCCCTCGGGCAGCTGCTCAGGAAGCCGAGGCGCAGTGA	1982
Qy	1021	GCCTCTCACCGAGGCCCTCGCTCGCATCTCCGAGGGTGCCTGGCCGACCCGCGCGC	1080
Db	1983	GCCTCTCACCGAGGCCCTCGCTCGCATCTCCGAGGGTGCCTGGCCGACCCGCGCGC	2042
Qy	1081	CCAGGGTGTGCTGCCCGGACACAGGTCATCCGAGGCTGTTCACACAGGCGGGTGA	1140
Db	2043	CCAGGGTGTGCTGCCCGGACACAGGTCATCCGAGGCTGTTCACACAGGCGGGTGA	2102
Qy	1141	GAGCCCTGAGCACGGGGCCCTGGCCCGGGGAGCAGCGGCACCCGGCCCGCGCGGCA	1200
Db	2103	GAGCCCTGAGCACGGGGCCCTGGCCCGGGGAGCAGCGGCACCCGGCCCGCGCGGCA	2162
Qy	1201	CCTCTGAAGGGCGGGTACATTGCGGGGGCTGCTCCAGGCTGCGCGAGCCACTGATGA	1260
Db	2163	CCTCTGAAGGGCGGGTACATTGCGGGGGCTGCTCCAGGCTGCGCGAGCCACTGATGA	2222
Qy	1261	GCACCGGTGTGAGAGAGAGGCGCGCAGGGAGGAGCAGCCACTCTCTGGGCCAAGC	1320
Db	2223	GCACCGGTGTGAGAGAGAGGCGCGCAGGGAGGAGCAGCCACTCTCTGGGCCAAGC	2282
Qy	1321	CCCTCATTCGAGACTGCGCTCGGCTGCTCTGGCACCCACTTTGGCCCTCTGGGCA	1380
Db	2283	CCCTCATTCGAGACTGCGCTCGGCTGCTCTGGCACCCACTTTGGCCCTCTGGGCA	2342
Qy	1381	CAGCCACTCTCCCTGGAACATGACTCTCCGAGCACCCCGCCCTCTCTCGAGGCTCGCG	1440
Db	2343	CAGCCACTCTCCCTGGAACATGACTCTCCGAGCACCCCGCCCTCTCTCGAGGCTCGCG	2402
Qy	1441	TGAGGCAAGGCACTGCTTCAGCCCTCGGGGGGGCCCTATCAGGGACATGGGGCA	1500
Db	2403	TGAGGCAAGGCACTGCTTCAGCCCTCGGGGGGGCCCTATCAGGGACATGGGGCA	2462
Qy	1501	CCCTCAGGGCTCCAGCAGCTTCCATCCACTGTGGCCACCCAGGACCTGCTCAGCCAGA	1560
Db	2463	CCCTCAGGGCTCCAGCAGCTTCCATCCACTGTGGCCACCCAGGACCTGCTCAGCCAGA	2522
Qy	1561	GAGGCCATTCGCCGACAGCCCTTGGGGGACGCCAGCCCTTTCTGCCACCCCAAGCAGGG	1620
Db	2523	GAGGCCATTCGCCGACAGCCCTTGGGGGACGCCAGCCCTTTCTGCCACCCCAAGCAGGG	2582
Qy	1621	TTCTGCCCCCGAGGGGTGACGCCCCACCCAGCAGTTGCCCATGCCCTCTCTGGCTC	1680
Db	2583	TTCTGCCCCCGAGGGGTGACGCCCCACCCAGCAGTTGCCCATGCCCTCTCTGGCTC	2642
Qy	1681	CTTCCCTCCAGGATCTTGCAGAGGGCCCTTAGTACCTCTCAGCCCTCTTGGGACA	1740
Db	2643	CTTCCCTCCAGGATCTTGCAGAGGGCCCTTAGTACCTCTCAGCCCTCTTGGGACA	2702
Qy	1741	GCCCGAGGACCCCTGCGCCCTGCAAGAGAGCCCCCACTTGGACTCTAAGATGGGGCC	1800
Db	2703	GCCCGAGGACCCCTGCGCCCTGCAAGAGAGCCCCCACTTGGACTCTAAGATGGGGCC	2762

QY 1801 TGGAGACATCTCTCTTCTGGGAGGCCAAAACCCGGCCCTCTGCAGTTTCCCGAGGTCAGC 1860
Db 2763 TGGAGACATCTCTCTTCTGGGAGGCCAAAACCCGGCCCTCTGCAGTTTCCCGAGGTCAGC 2822
QY 1861 CTCCAGGGAGAGCTCTTCCCAAGTAGCTCTCTCAAGGTGGGCTCTCTCCAGAGTGGGCAC 1920
Db 2823 CTCCAGGGAGAGCTCTTCCCAAGTAGCTCTCTCAAGGTGGGCTCTCTCCAGAGTGGGCAC 2882
QY 1921 AGAGCTGGCCCTCTCGATGCGGAGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 1980
Db 2883 AGAGCTGGCCCTCTCGATGCGGAGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 2942
QY 1981 CTCCACACCATCTTGACAGGGCTCAGGAACAGGCGACATGCGCAAGTTCTCCCTGGG 2040
Db 2943 CTCCACACCATCTTGACAGGGCTCAGGAACAGGCGACATGCGCAAGTTCTCCCTGGG 3002
QY 2041 TGGTCCGGGGGCTACGAGGCTGGCTATGGCACCTTTGGTGGAGATGC 2100
Db 3003 TGGTCCGGGGGCTACGAGGCTGGCTATGGCACCTTTGGTGGAGATGC 3062
QY 2101 AGGGGCGATGCTGGGCGAGGGGCCATGTGGGCCAGGATAGCTGGGCTGTGCCAGTC 2160
Db 3063 AGGGGCGATGCTGGGCGAGGGGCCATGTGGGCCAGGATAGCTGGGCTGTGCCAGTC 3122
QY 2161 GGAGGAGGAGGAGGAGGAGGGCCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2220
Db 3123 GGAGGAGGAGGAGGAGGAGGGCCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 3182
QY 2221 CAGGCTGAGAGCCACTGCCCCAGCTCAGTGCAGAGGCTGTGCTGAGGTGCGGAGGCG 2280
Db 3183 CAGGCTGAGAGCCACTGCCCCAGCTCAGTGCAGAGGCTGTGCTGAGGTGCGGAGGCG 3242
QY 2281 TCCCAACAGGAGCTCTCAGAGGCCACCCATGAGGAGGACATCGGGAGGCTCCCTGGT 2340
Db 3243 TCCCAACAGGAGCTCTCAGAGGCCACCCATGAGGAGGACATCGGGAGGCTCCCTGGT 3302
QY 2341 CGAGATCCGGAGCTGTGAGTGATGCGGAGGCGCGCACAAATATCCCTGGACATTTTC 2400
Db 3303 CGAGATCCGGAGCTGTGAGTGATGCGGAGGCGCGCACAAATATCCCTGGACATTTTC 3362
QY 2401 CGAGTGACCCCGCTACCTCACTCTCAGACTGTGACATATCAAGTACCTCCCAT 2460
Db 3363 CGAGTGACCCCGCTACCTCACTCTCAGACTGTGACATATCAAGTACCTCCCAT 3422
QY 2461 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCAT 2520
Db 3423 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCAT 3482
QY 2521 GGCTGAGGAGAGCTGGCCGAGTTCCCGAGGCCCAAGTGGCCCTTGGCCAGGTGAATGGG 2580
Db 3483 GGCTGAGGAGAGCTGGCCGAGTTCCCGAGGCCCAAGTGGCCCTTGGCCAGGTGAATGGG 3542
QY 2581 CCCCAACAGGCTTGAGATCACAGAGGAGTCAAGAGTGTGACCGGCTGTGGGAGGAG 2640
Db 3543 CCCCAACAGGCTTGAGATCACAGAGGAGTCAAGAGTGTGACCGGCTGTGGGAGGAG 3602
QY 2641 GGCTGCCGTGGCAGGAGCGCAAGTGTCTCCCGCTCAGCAGAGCTCTTCCACTTCCC 2700
Db 3603 GGCTGCCGTGGCAGGAGCGCAAGTGTCTCCCGCTCAGCAGAGCTCTTCCACTTCCC 3662
QY 2701 TGGGAGCAGCTGCGGCTGGATGAGCTTCAGAGCTGGGCTGGGTGAGAGGTGAAGGC 2760
Db 3663 TGGGAGCAGCTGCGGCTGGATGAGCTTCAGAGCTGGGCTGGGTGAGAGGTGAAGGC 3722
QY 2761 CTCGCTGGAGCACATCTCCCGATCTCAAGGGGAGGCCCGGAAGTCTGGAGAGGAGGG 2820
Db 3723 CTCGCTGGAGCACATCTCCCGATCTCAAGGGGAGGCCCGGAAGTCTGGAGAGGAGGG 3782
QY 2821 GCGCCCGAGGAGGAGCGAGGCTTGTCTCTCCGCTCTCAGGTCTGAAGAGCTGGGA 2880
Db 3783 GCGCCCGAGGAGGAGCGAGGCTTGTCTCTCCGCTCTCAGGTCTGAAGAGCTGGGA 3842

QY 2881 CCGAGCGCCGACATTTCTTAGGGAGCTCTCAGATGAGACTGTGCTCTGGGCCAGTCAGT 2940
Db 3843 CCGAGCGCCGACATTTCTTAGGGAGCTCTCAGATGAGACTGTGCTCTGGGCCAGTCAGT 3902
QY 2941 GACACTGGCTGCCAGAGTGTGAGCCAGCCAGCTGCCAGGCCACCTGGGACCAAGACGG 3000
Db 3903 GACACTGGCTGCCAGAGTGTGAGCCAGCCAGCTGCCAGGCCACCTGGGACCAAGACGG 3962
QY 3001 A 3001
Db 3963 A 3963
RESULT 10
AX039409
LOCUS
DEFINITION
Sequence 2 from Patent WO0063381.
ACCESSION
AX039409
VERSION
AX039409.1 GI:11229476
KEYWORDS
Homo sapiens (human)
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Zeng W., Stanton, L. and Kong, H.
Mammalian protein with putative function in signal transduction
Patent: WO 0063381-A 2 26-OCT-2000;
SCIOS INC. (US)
FEATURES
Location/Qualifiers
1..4175
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
<1..4056
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC16624.1"
/db_xref="GI:11229477"
/translation="GVQFSQYGSPEFVSPETIQNPVSEASDIWANGVTSYLSLTCS
PFAGESDRATLLNLEGRVSWSPMAHLSEDAKOFIKATLQRAPOARPSAQCPSLHP
WFLGNPABEAHFINTKQLFLARSRWQSLMSYKILVMRSIPELLRGPDPSPSLG
VARHLCDRTGSSSSSSSDNELAPFAKSLPPSPVTHSPLLHPRGFRGALAPGSR
AASRSRTAPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP
HPARRHLKGGVIGALPLREPLMEHVELEEAREREOATLLAKAPSFETALRLPA
SHTHLAPGSHSLHDSPTPRPSEACEAQLPSAPSGGAPIDRMGHPQSGKQLPS
TGHTGTAQPERPSPDPSWQGPAPCPKPKQGSAPQEGCSPHPAVAPCPGSPSPGCK
EAPLPSPFLGPQAPAPAKAPPLDSKMGPGDISLPRPKPGPCSPGSAQSAS
SQVSLRVSSQVTEPGSLDAEGWTQEAEDLSPTTLQRPQOATWRFSLGRRG
GYAGVAGYTFAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
AESPLPOVSARPVGVGRAPTRSSPEPTPWEIDIGOVSLVQIRDLGDAEADTISLDI
SEVDPAYNLSLDYIKYLPFEFMI FRKVPKSAQPEPPSPMAEELAEFPETWPFPG
ELGPHAGLEITEESDVKALLAEAVGRRKRWSSPSRSLFHPGRHLPLDEPAELGLR
ERKVASVEHISRLKRGPEKEGPRKRGSLFRSLGSLKSWDRAPTFRLSDELT
VYLGQSVTLACQVAPAAQWAKSDGAPLESSRVLISATLNKQLLTILVYVADL
GYTCSVNLGTVTTVLRKAERPSGPCPDIGEVYADGVLVVKPVSVPVYI
VOCLEGGSWTTLASDI PDCCVLTSTKSRGTYTFRCTACVSKAGMPYSPSEVILG
GSHLSREESQGRSAQPLPSTKTAFTQIORGFVSUVRQCKEASGRALAAKIIPY
HPKDTAVLREYALKGRHPLAQHAYLSFRHLVLELCSFELLPCLAEARASY
SESEVKYLMQMSATQYLNQHIUHLDRSENMIITNYNLLKLVLDLNGAQLSDEKV
LPSDFKDYLETMAPELLEGGAVPTDIAWGTAFIMLSAEYFVSEGGADLQRL
RKGLVRLRCVAGLSGAVAFRLSTLCAQPMGRPCASSCLQCPMLTEBGPACSRPAPY
TPTARLRFVRNREKRALLYKRHLAQVR"
ORIGIN
Query Match 88.1%; Score 2643.4; DB 6; Length 4175;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2644; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 357 TGCAGTTACGCGATACGCTCCCTGAGTTTCGTTCTCCCGAGATCATCCAGCAGAAC 416
Db 5 TGCAGTTACGCGATACGCTCCCTGAGTTTCGTTCTCCCGAGATCATCCAGCAGAAC 64

Qy	417	CTGTGAGCGAAGCCCTCCGACATTTGGGGGCATGGGTGTCACTCTCTA	CCTCAGCCTGACCT	476
Db	65	CTGTGAGCGAAGCCCTCCGACATTTGGGGCCATGGGTGTCACTCTCTA	CCTCAGCCTGACCT	124
Qy	477	GCTCATCCCATTTGCGGGAGAGTGACCGTGCACCCCTCTCTGAAAGTCTCT	CGAGGGGC	536
Db	125	GCTCATCCCATTTGCGGGAGAGTGACCGTGCACCCCTCTCTGAAAGTCTCT	CGAGGGGC	184
Qy	537	GCGTGTCTATGGAGCAGGCCCATGTGCTGCCACCTCAGCGAAGCGCCAAAGACTT	CATCA	596
Db	185	GCGTGTCTATGGAGCAGGCCCATGTGCTGCCACCTCAGCGAAGCGCCAAAGACTT	CATCA	244
Qy	597	AGGCTAGCTGCAGAGACCCCTCAGCGCCGGCCCTAGTGGCGGCCCATGTGCTCT	CCACC	656
Db	245	AGGCTAGCTGCAGAGACCCCTCAGCGCCGGCCCTAGTGGCGGCCCATGTGCTCT	CCACC	304
Qy	657	CCTGGTTCCTGAATCCATGCTCGGAGAGGCCCACTTCATCAACACCAAGCAGCTCA	716	
Db	305	CCTGGTTCCTGAATCCATGCTCGGAGAGGCCCACTTCATCAACACCAAGCAGCTCA	364	
Qy	717	AGTTCTCTCGGCCCGAAGTCGCTGGCAGCGGTTCCCTGATGAGCTACAAGTCCATCCT	CGG	776
Db	365	AGTTCTCTCGGCCCGAAGTCGCTGGCAGCGGTTCCCTGATGAGCTACAAGTCCATCCT	CGG	424
Qy	777	TGATGCGCTCATCCCTGATGCTGTGCGGGGCCCAACCGACAGCCCTCCTCTGGGGTAG	836	
Db	425	TGATGCGCTCATCCCTGATGCTGTGCGGGGCCCAACCGACAGCCCTCCTCTGGGGTAG	484	
Qy	837	CCGGGCACTCTGACAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTGACAAAG	896	
Db	485	CCGGGCACTCTGACAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTGACAAAG	544	
Qy	897	AGCTCGGCCCATTTGCCCGGGCTAAAGTCACTGCCACCGTCCCGGGTGACACACTCACAC	956	
Db	545	AGCTCGGCCCATTTGCCCGGGCTAAAGTCACTGCCACCGTCCCGGGTGACACACTCACAC	604	
Qy	957	TGCTGCACCCCGGGGCTTCTGCGGCCCTTGGCGACCTGCCGTGAGGAAGCCGAGGCCA	1016	
Db	605	TGCTGCACCCCGGGGCTTCTGCGGCCCTTGGCGACCTGCCGTGAGGAAGCCGAGGCCA	664	
Qy	1017	GTGAGCGCTCCACGAGGCCCAAGCTCGGCTGCTATCTCCGAGGGTGC	CGGGCCACCG	1076
Db	665	GTGAGCGCTCCACGAGGCCCAAGCTCGGCTGCTATCTCCGAGGGTGC	CGGGCCACCG	724
Qy	1077	CCGCCCAGGGCTGGTGCCCGGCACAGCGTCACTCCGACGCTGTTCTACACACAGCGCG	1136	
Db	725	CCGCCCAGGGCTGGTGCCCGGCACAGCGTCACTCCGACGCTGTTCTACACACAGCGCG	784	
Qy	1137	GTGAGAGCCCTGAGCACGGGGCCCTGGCCCCGGGGAGCAGCGGCACCCCGCCGGCGCGC	1196	
Db	785	GTGAGAGCCCTGAGCACGGGGCCCTGGCCCCGGGGAGCAGCGGCACCCCGCCGGCGCGC	844	
Qy	1197	GGCACTGTGTAAGGGGGGTACATTTGGGGGGGGCTGCCAGGCTCTGGGAGGCCACTGA	1256	
Db	845	GGCACTGTGTAAGGGGGGGGTACATTTGGGGGGGGCTGCCAGGCTCTGGGAGGCCACTGA	904	
Qy	1257	TGGAGCACCGGTCTGTGAGAGAGAGCGCGCACGGAGGAGCAGGCCACCTCTCTGGGCCA	1316	
Db	905	TGGAGCACCGGTCTGTGAGAGAGAGCGCGCACGGAGGAGCAGGCCACCTCTCTGGGCCA	964	
Qy	1317	AAGCCCCCTCATTCGAGACTGCCCCCTCGGGTGTCTCTGCGCACCCACTTGGCCCCCTG	1376	
Db	965	AAGCCCCCTCATTCGAGACTGCCCCCTCGGGTGTCTCTGCGCACCCACTTGGCCCCCTG	1024	
Qy	1377	GCACAGCCACTCCCTGGAAATGACTCTCGAGGACCGCCCGGCCCTCTCTCGAGGCGCT	1436	
Db	1025	GCACAGCCACTCCCTGGAAATGACTCTCGAGGACCGCCCGGCCCTCTCTCGAGGCGCT	1084	
Qy	1437	CGCGTAGGGCACAGCGACTGCTTTACGCCCCCTCGGGGGGGCCCCCTATCAGGGACATGG	1496	
Db	1085	CGCGTAGGGCACAGCGACTGCTTTACGCCCCCTCGGGGGGGCCCCCTATCAGGGACATGG	1144	
Qy	1497	GGCACCTCTCAGGGCTCCAAGCAGCTTCATCCACTGTGTGGCCACCCAGGACATGCTCAGC	1556	

Db	1145	GGCACCTCAGGGCTCCAAAGACGCTTCCATCCACTGGTGGCCACCCAGGCACTGCCTCAGC	1204
Qy	1557	CAGAGAGCCATCCCCGGACAGAGCCCTTGGGGGAGCCAGCCCTTCTGTGCACCCCAAGC	1616
Db	1205	CAGAGAGCCATCCCCGGACAGAGCCCTTGGGGGAGCCAGCCCTTCTGTGCACCCCAAGC	1264
Qy	1617	AGGGTTCTGCCCCCAAGAGGGGTGACAGCCCCCAACCCAGAGAGTTGGCCCCATGCCCTCTG	1676
Db	1265	AGGGTTCTGCCCCCAAGAGGGGTGACAGCCCCCAACCCAGAGAGTTGGCCCCATGCCCTCTG	1324
Qy	1677	GCTCTCTCCCTCAGAGATCTTGCAAGAGAGGCCCCCTTAGTACCCTCAAGGCCCTTCTTGG	1736
Db	1325	GCTCTCTCCCTCAGAGATCTTGCAAGAGAGGCCCCCTTAGTACCCTCAAGGCCCTTCTTGG	1384
Qy	1737	GACAGCCCCAGGCAACCCCTGCCCTGCCAAAGCAAGCCCCCATTTGGACTCTTAAGATGG	1796
Db	1385	GACAGCCCCAGGCAACCCCTGCCCTGCCAAAGCAAGCCCCCATTTGGACTCTTAAGATGG	1444
Qy	1797	GGGCTTGAGACATCTCTCTTCTTGGAGGCCAAAACCCGGCCCCCTCGAGTTCCCCAGGGT	1856
Db	1445	GGGCTTGAGACATCTCTCTTCTTGGAGGCCAAAACCCGGCCCCCTCGAGTTCCCCAGGGT	1504
Qy	1857	CAGGCTCCAGGAGAGCTCTTCCCAAGTAGCTCCCTCAGGGTGGGCTCCTCCAGGTTGG	1916
Db	1505	CAGGCTCCAGGAGAGCTCTTCCCAAGTAGCTCCCTCAGGGTGGGCTCCTCCAGGTTGG	1564
Qy	1917	GCACAGAGCTTGGCCCTCCCTCGATCGGAGGGCTGGACCCAGGAGGCTGAGGATCTGT	1976
Db	1565	GCACAGAGCTTGGCCCTCCCTCGATCGGAGGGCTGGACCCAGGAGGCTGAGGATCTGT	1624
Qy	1977	CCGACTCCACACCACTTGCAGCGGCTCAGGAACAGGACCATGCGCAAGTTCTCCC	2036
Db	1625	CCGACTCCACACCACTTGCAGCGGCTCAGGAACAGGACCATGCGCAAGTTCTCCC	1684
Qy	2037	TGGGTGTTCGGGGGCTAGCGAAGGCGTGGCTGGCTATGGCACCTTTGGCTTTGGTGGAG	2096
Db	1685	TGGGTGTTCGGGGGCTAGCGAAGGCGTGGCTGGCTATGGCACCTTTGGCTTTGGTGGAG	1744
Qy	2097	ATGCAGGGGCACTGTGGGGCAGGGGCCCATGTGGGCAGGATAGCTGGGCTGTGTCCC	2156
Db	1745	ATGCAGGGGCACTGTGGGGCAGGGGCCCATGTGGGCAGGATAGCTGGGCTGTGTCCC	1804
Qy	2157	AGTCGAGGAGGAGGAGCAGGAGGCCAGGGCTGAGTCCCAGTCGGAGGAGCAGCAGG	2216
Db	1805	AGTCGAGGAGGAGGAGCAGGAGGCCAGGGCTGAGTCCCAGTCGGAGGAGCAGCAGG	1864
Qy	2217	AGGCAAGGGCTGAGAGCCCACTGCCCAAGTCTAGTGCAGGCCCTGTGCTGAGGTTGGCA	2276
Db	1865	AGGCAAGGGCTGAGAGCCCACTGCCCAAGTCTAGTGCAGGCCCTGTGCTGAGGTTGGCA	1924
Qy	2277	GGGCTCCACAGGAGCTCTCAGAGGCCACCCATGGGAGGACATCGGSCAGGCTCTCCC	2336
Db	1925	GGGCTCCACAGGAGCTCTCAGAGGCCACCCATGGGAGGACATCGGSCAGGCTCTCCC	1984
Qy	2337	TGTTGCAGATCCGGGACCTGTCAAGTGTATGCGGAGGCGGCCGACACAATATCCCTGGACA	2396
Db	1985	TGTTGCAGATCCGGGACCTGTCAAGTGTATGCGGAGGCGGCCGACACAATATCCCTGGACA	2044
Qy	2397	TTTCCGAGGTGGACCCCGCCTAAGCTCAAGCTGTGAGAGTATCAAGTACCTTCC	2456
Db	2045	TTTCCGAGGTGGACCCCGCCTAAGCTCAAGCTGTGAGAGTATCAAGTACCTTCC	2104
Qy	2457	CATTTCAGATTATGATCTTCAGAGAAAGTCCCCAAGTCCGCTCAGGCAGAGCCGCCCTCCC	2516
Db	2105	CATTTCAGATTATGATCTTCAGAGAAAGTCCCCAAGTCCGCTCAGGCAGAGCCGCCCTCCC	2164
Qy	2517	CCATGGCTGAGGAGGAGCTGGCCGAGTTCCCGGAGGCCACGTTGGCCCTGCGCCAGGTGAAC	2576
Db	2165	CCATGGCTGAGGAGGAGCTGGCCGAGTTCCCGGAGGCCACGTTGGCCCTGCGCCAGGTGAAC	2224
Qy	2577	TGGGCCCCCCACGCAAGGCTGGAGATCAAGAGAGTCAAGAGATGTGGACGCGCTGTCTGG	2636

1041 QY CTCGCGCTGCATCTCCGAGGGTCCGGGCGCACGGGCGCCACAGGGCTGCGTGCCCGGC 1100
601 Db CTCGCGCTGCATCTCCGAGGGTGCCTGGGCGCACGGGCGCCACAGGGCTGCGTGCCCGGC 660
1101 QY ACAGCGTCAATCCGACGACCTGTTCTACCAACAGGCGGGTGAGAGCCCTCAGACACGGGGGCC 1160
661 Db ACAGCGTCAATCCGACGACCTGTTCTACCAACAGGCGGGTGAGAGCCCTCAGACACGGGGGCC 720
1161 QY TGGCCCGCGGGAGCAGGGCGGCACCGGCGCCCGCGCGCGCACTGCTGTAAGCGCGGGTACA 1220
721 Db TGGCCCGCGGGAGCAGGGCGGCACCGGCGCCCGCGCGCGCACTGCTGTAAGCGCGGGTACA 780
1221 QY TTGGCGGGGCGCTTSCCAGGCGCTGGCGAGCCACTGATGGAGCACCGCTGCTGGAGAGG 1280
781 Db TTGGCGGGGCGCTTSCCAGGCGCTGGCGAGCCACTGATGGAGCACCGCTGCTGGAGAGG 840
1281 QY AGGCGCGCAGGGAGGAGCAGGCGCACCTCTCTGGCCAAAGCCCGCTCATTTCCAGAGACTGCC 1340
841 Db AGGCGCGCAGGGAGGAGCAGGCGCACCTCTCTGGCCAAAGCCCGCTCATTTCCAGAGACTGCC 900
1341 QY TCCGCGCTGCTGCTCTGGCAACCACTTTGGGCCCTTGCGCCACAGGCCACTTCCCTTGAACATG 1400
901 Db TCCGCGCTGCTGCTCTGGCAACCACTTTGGGCCCTTGCGCCACAGGCCACTTCCCTTGAACATG 960
1401 QY ACTCTCCGAGCAGCCCCCGCCCTCTCCGAGGCGCTGCGGTGAGGCACAGCGACTGCCCTT 1460
961 Db ACTCTCCGAGCAGCCCCCGCCCTCTCCGAGGCGCTGCGGTGAGGCACAGCGACTGCCCTT 1020
1461 QY CAGCGCCCTCTCGGGGGGGCCCTATCAGGGACATGGGGCACCTCAGGGCTCCAAAGCAGC 1520
1021 Db CAGCGCCCTCTCGGGGGGGCCCTATCAGGGACATGGGGCACCTCAGGGCTCCAAAGCAGC 1080
1521 QY TTCCATCACTGGTGGCCACCCAGGCACTGTACGCCAGAGAGGCGCATCCCGGACAGCC 1580
1081 Db TTCCATCACTGGTGGCCACCCAGGCACTGTACGCCAGAGAGGCGCATCCCGGACAGCC 1140
1581 QY CTTGGGGGCGCAGCGCCCTTCTGCGCACCCAGAGGCTTCTGCGCCCGCAGAGGGCT 1640
1141 Db CTTGGGGGCGCAGCGCCCTTCTGCGCACCCAGAGGCTTCTGCGCCCGCAGAGGGCT 1200
1641 QY GCAGCGCCCAACCCAGCAGTGTGCGCCCTCTGCGCTCTCTCCCTCCAGGATCTTGCA 1700
1201 Db GCAGCGCCCAACCCAGCAGTGTGCGCCCTCTGCGCTCTCTCCCTCCAGGATCTTGCA 1260
1701 QY AAGAGGCGCCCTTAGTACCTCAAGCCCTTCTTTGGGACAGCCCGCAGGCAACCCCTGCCC 1760
1261 Db AAGAGGCGCCCTTAGTACCTCAAGCCCTTCTTTGGGACAGCCCGCAGGCAACCCCTGCCC 1320
1761 QY CTGCGCAAGCAAGCCCCCATTTGACTCTAAGATGGGGCGCTGGAGACATCTCTTCTCTG 1820
1321 Db CTGCGCAAGCAAGCCCCCATTTGACTCTAAGATGGGGCGCTGGAGACATCTCTTCTCTG 1380
1821 QY GGAGGCCAAAACCGGGCCCTGCACTTCCCGAGGCTCAGGCTTCCCGAGCGAGCTCTTCCC 1880
1381 Db GGAGGCCAAAACCGGGCCCTGCACTTCCCGAGGCTCAGGCTTCCCGAGCGAGCTCTTCCC 1440
1881 QY AAGTGAAGTCTCCCTCAGGGTGGGCTCTCTCCAGGTGGGACAGAGCTTGGCCCTCCCTGG 1940
1441 Db AAGTGAAGTCTCCCTCAGGGTGGGCTCTCTCCAGGTGGGACAGAGCTTGGCCCTCCCTGG 1500
1941 QY ATGGGGAGGCTGGAACCCAGAGGCTGAGGATCTGTCCGACTCCACACCCACCTTGCAGC 2000
1501 Db ATGGGGAGGCTGGAACCCAGAGGCTGAGGATCTGTCCGACTCCACACCCACCTTGCAGC 1560
2001 QY GGCCTCAGGAACAGCGACCATGCGCAAGTTCTCCCTGGGTGGTCCGGGGGCTACGCAG 2060
1561 Db GGCCTCAGGAACAGCGACCATGCGCAAGTTCTCCCTGGGTGGTCCGGGGGCTACGCAG 1620
2061 QY CGCTGGCTGGCTATGGCACTTGTGCTTTGGTGGAGATGCGAGGGGCGCATGCTGGGGCAGG 2120
1621 Db CGCTGGCTGGCTATGGCACTTGTGCTTTGGTGGAGATGCGAGGGGCGCATGCTGGGGCAGG 1680
2121 QY GGCCCATGTGGGCCAGGATAGCTGGGCTGTGTCCAGTCCGAGTGGAGGAGGAGCAGGAGG 2180

1581 Db GGCCCATGTGGGCCAGGATAGCTTGGGCTGTGTCCAGTCCGAGGAGGAGGAGCAGGAGG 1740
2181 QY AGGCGAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGCGCAGGGCTGAGAGCCCACTGC 2240
1741 Db AGGCGAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGCGCAGGGCTGAGAGCCCACTGC 1800
2241 QY CCAGAGTCAAGTCAAGGCGCTGTGCTGAGTCCGAGGAGGCTCCCAACAGGAGCTCTCCAG 2300
1801 Db CCAGAGTCAAGTCAAGGCGCTGTGCTGAGTCCGAGGAGGCTCCCAACAGGAGCTCTCCAG 1860
2301 QY AGCCACACCCATGGAGGAGCATCGGCGAGGCTCTCCCTGGTGCAGATCCGAGACCTGTGAG 2360
1861 Db AGCCACACCCATGGAGGAGCATCGGCGAGGCTCTCCCTGGTGCAGATCCGAGACCTGTGAG 1920
2361 QY GTGATGCGGAGGCGGCGCAGACAAATATCCCTTGGACATTTCCGAGGTGAGACCCCGCTTACC 2420
1921 Db GTGATGCGGAGGCGGCGCAGACAAATATCCCTTGGACATTTCCGAGGTGAGACCCCGCTTACC 1980
2421 QY TCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAATTCGAGTTTATGATCTTCAGGA 2480
1981 Db TCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAATTCGAGTTTATGATCTTCAGGA 2040
2481 QY AAGTCCCAAGTCCGCTCAGCAGAGCGCGCTCCCGCATCGCTGAGGAGGAGCTGGCGG 2540
2041 Db AAGTCCCAAGTCCGCTCAGCAGAGCGCGCTCCCGCATCGCTGAGGAGGAGCTGGCGG 2100
2541 QY AGTTCCCGGAGGCCACCTGGGCGCTGAGGAGGTAACTGGGCGCCCAACAGGAGCTTGGAGA 2600
2101 Db AGTTCCCGGAGGCCACCTGGGCGCTGAGGAGGTAACTGGGCGCCCAACAGGAGCTTGGAGA 2160
2601 QY TCACAGAGGAGTCAAGAGATGTGACGCGCTGTGCGACGAGGCTGCGGTGGGCGAGGAGC 2660
2161 Db TCACAGAGGAGTCAAGAGATGTGACGCGCTGTGCGACGAGGCTGCGGTGGGCGAGGAGC 2220
2661 QY GCAAGTGTCTCCGCTCAGCAGAGCTTCTTCCACTTCCCTGGGAGGAGCAGCTTGGCGCTGG 2720
2221 Db GCAAGTGTCTCCGCTCAGCAGAGCTTCTTCCACTTCCCTGGGAGGAGCAGCTTGGCGCTGG 2280
2721 QY ATGAGCTTGCAGAGCTGGGCGCTGCTGAGAGGTGAAGGCTTCCGTGGAGCAGCATCTCCC 2780
2281 Db ATGAGCTTGCAGAGCTGGGCGCTGCTGAGAGGTGAAGGCTTCCGTGGAGCAGCATCTCCC 2340
2781 QY GGATCTTGAAGGCGAGCGCGGAGGCTTGGAGAGGAGGAGGCGCGCCCGCAGGAGAGCCAG 2840
2341 Db GGATCTTGAAGGCGAGCGCGGAGGCTTGGAGAGGAGGAGGCGCGCGCCCGCAGGAGAGCCAG 2400
2841 QY GCCTTGTCTTCCGCTCTCAGGCTCTGAGAGCTGAGAGCTGGGACCGGCGGCGCATTTCTTAA 2900
2401 Db GCCTTGTCTTCCGCTCTCAGGCTCTGAGAGCTGAGAGCTGGGACCGGCGGCGCATTTCTTAA 2460
2901 QY GGGAGCTTCAAGATGAGATGTGTGCTTCCCTGGGCGCAGTCAAGTGCACATGCGCTGCCAGGTGT 2960
2461 Db GGGAGCTTCAAGATGAGATGTGTGCTTCCCTGGGCGCAGTCAAGTGCACATGCGCTGCCAGGTGT 2520
2961 QY CAGCCCGCAGCAGCTGCGCCAGGCGCACCTGGAGCAAGACGGA 3001
2521 Db CAGCCCGCAGCAGCTGCGCCAGGCGCACCTGGAGCAAGACGGA 2561

RESULT: 12
AL670729
LOCUS
DEFINITION Human DNA sequence from clone RP11-245P10 on chromosome 1, complete sequence.
ACCESSION AL670729
VERSION AL670729.19 GI:22316158
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 62164)

AUTHORS TITLE JOURNAL

Van Hellmond, Z.
Direct Submission
Submitted (15-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 19, 2002 this sequence version replaced gi:21727388.

COMMENT

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
RP11-245P10 is from the library RPci-11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6.

FEATURES

Location/Qualifiers
1..62164
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-245P10"
/clone_lib="RPci-11.1"

ORIGIN

Query Match 69.0%; Score 2070.4; DB 9; Length 62164;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2077; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 719 TTCTCTCTGGCCGGAAGTCGTGGCAGCGTTCCCTGATGAGCTACAAAGTCATCTCGGTG 778
Db 24359 TGCCTGTGCCCCCATCGGTTCCAGGTTCCCTGATGAGTACAAAGTCATCTCGGTG 24418
Qy 779 ATGGCTTCATCTCTGAGCTGCTCGGGGGCCACCCGACAGCCCTCCCTCGGGTAGCC 838
Db 24419 ATGGCTTCATCTCTGAGCTGCTCGGGGGCCACCCGACAGCCCTCCCTCGGGTAGCC 24478
Qy 839 CGGCACCTCTGACGGACACTGGTGGCTCTCCAGTTCTCTCCCTCTCTGACACAGAG 898
Db 24479 CGGCACCTCTGACGGACACTGGTGGCTCTCTCCAGTTCTCTCTCTCTGACACAGAG 24538
Qy 899 CTGCCCCATTTGCCGGGCTAAGTCACTGCACCTCTCCCGGTGACACACTCACCAGTG 958
Db 24539 CTGCCCCATTTGCCGGGCTAAGTCACTGCACCTCTCCCGGTGACACACTCACCAGTG 24598
Qy 959 CTGCACCCCGGGGCTTCTCGGCGCTCTCGGCCAGCTGCTGAGGAAGCCGAGGCCAGT 1018
Db 24599 CTGCACCCCGGGGCTTCTCGGCGCTCTCGGCCAGCTGCTGAGGAAGCCGAGGCCAGT 24658
Qy 1019 GAGCGCTCCACGAGGCCCCAGCTCCGCGCTGCATCTCCCGAGGGTGGCCGGCCACCGCC 1078

Db	24659	GAGCGCTCCACCGAGGCCCCAGCTCCGCTGCATCTCCCGAGGGTGC CGGGCCACCGGCC	24718
Qy	1079	GCCGAGGCTGCGTGGCCCCCGGACAGCGTCTATCCGAGCCTGTGTTACCAACAGCGGGGT	1138
Db	24719	GCCGAGGCTGCGTGGCCCCCGGACAGCGTCTATCCGAGCCTGTGTTACCAACAGCGGGGT	24778
Qy	1139	GAGAGCCCTGAGCACGGGGGCTGGCCCCGGGAGAGCGGACACCGCGGCCGCGCGG	1198
Db	24779	GAGAGCCCTGAGCACGGGGGCTGGCCCCGGGAGAGCGGACACCGCGGCCGCGCGG	24838
Qy	1199	CACCTGTGAAGGCGGTACATTTGCGGGGCGCTGCCAGCGCTTGC CGAGCCACTGATG	1258
Db	24839	CACCTGTGAAGGCGGTACATTTGCGGGGCGCTGCCAGCGCTTGC CGAGCCACTGATG	24898
Qy	1259	GAGCACCGCTGCTGGAGGAGGCGCCGACAGGAGAGAGAGCGGACACCTCTCTGGCCAAA	1318
Db	24899	GAGCACCGCTGCTGGAGGAGGCGCCGACAGGAGAGAGAGCGGACACCTCTCTGGCCAAA	24958
Qy	1319	GCCCCCTCATTTGAGAGCTGCCCTCCGGCTGCTGCCCTCTGGCACCCACTTTGGCCCTGGC	1378
Db	24959	GCCCCCTCATTTGAGAGCTGCCCTCCGGCTGCTGCCCTCTGGCACCCACTTTGGCCCTGGC	25018
Qy	1379	CACAGCCACTCTCTGGAAACATGACTCTCCGAGCACCCCGCCCTCTCTCGAGGCGCTGC	1438
Db	25019	CACAGCCACTCTCTGGAAACATGACTCTCCGAGCACCCCGCCCTCTCTCGAGGCGCTGC	25078
Qy	1439	GGTGAGGCACAGGAGCTGCCCTTCAGCCCCCTCCGGGGGGGCGCCCTATCAGGAGCATGGG	1498
Db	25079	GGTGAGGCACAGGAGCTGCCCTTCAGCCCCCTCCGGGGGGGCGCCCTATCAGGAGCATGGG	25138
Qy	1499	CACCTCAGGGCTCCAGGAGCTTCCATCCACTGTGGTGGCCACCCAGGCACTGTCTCAGCCA	1558
Db	25139	CACCTCAGGGCTCCAGGAGCTTCCATCCACTGTGGTGGCCACCCAGGCACTGTCTCAGCCA	25198
Qy	1559	GAGAGGCACTCCCGGACAGCCCTTGGGGGAGCGACAGCCCTTTCTGCGCACCCCAAGCAG	1618
Db	25199	GAGAGGCACTCCCGGACAGCCCTTGGGGGAGCGACAGCCCTTTCTGCGCACCCCAAGCAG	25258
Qy	1619	GGTTCTGCCCCCAGGAGGCTGCAGCCCCCACCAGCAGAGTTGCCCTCCCTCTCTGGC	1678
Db	25259	GGTTCTGCCCCCAGGAGGCTGCAGCCCCCACCAGCAGAGTTGCCCTCCCTCTCTGGC	25318
Qy	1679	TCCTTCTCTCAGGATCTTTCGAAAGAGGCCCTTAGTAGTACCTTCAAGCCCTCTCTGGGA	1738
Db	25319	TCCTTCTCTCAGGATCTTTCGAAAGAGGCCCTTAGTAGTACCTTCAAGCCCTCTCTGGGA	25378
Qy	1739	CAGCCCCAGGACCCCTGCCCCCTGCCAAAGCAAGCCCCCATTTGAGACTCTTAAGATGGG	1798
Db	25379	CAGCCCCAGGACCCCTGCCCCCTGCCAAAGCAAGCCCCCATTTGAGACTCTTAAGATGGG	25438
Qy	1799	CTTGAGAGCATCTCTCTCTGGAGGCCCAAAACCCGGCCCTTGCAGTTTCCCGAGGTC	1858
Db	25439	CTTGAGAGCATCTCTCTCTGGAGGCCCAAAACCCGGCCCTTGCAGTTTCCCGAGGTC	25498
Qy	1859	GCTTCCAGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCTCCAGAGTGGC	1918
Db	25499	GCTTCCAGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCTCCAGAGTGGC	25558
Qy	1919	ACAGAGCTTGGGCTCTCTCTGATGCGAGGGGTGACCCAGAGGCTGAGGATCTGTCC	1978
Db	25559	ACAGAGCTTGGGCTCTCTCTGATGCGAGGGGTGACCCAGAGGCTGAGGATCTGTCC	25618
Qy	1979	GACTCCACACCCACTTGACGGGCTCAGGAAACAGGCGACCATGGCAAGTTCTCCCTG	2038
Db	25619	GACTCCACACCCACTTGACGGGCTCAGGAAACAGGCGACCATGGCAAGTTCTCCCTG	2038
Qy	2039	GTTGTTGCGGGGCTTACAGCGGTGGCTGCTATGGCACTTTGCCCTTTGGTGGAGAT	2098
Db	25679	GTTGTTGCGGGGCTTACAGCGGTGGCTGCTATGGCACTTTGCCCTTTGGTGGAGAT	25738
Qy	2099	GCAGGGGCTCTCGGGCAGGGGCCCTATGTGGGCCAGGATAGCTGGGCTGTGTCCAG	2158

Db 25739 GCAGGGGCGATGCTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTCTCCAG 25798
Qy 2159 TCGAGGAGGAGGAGCAGGAGAGCCAGGGCTGAGTCCAGTTCGAGGAGCAGCAGAG 2218
Db 25799 TCAGAGGAGGAGGAGCAGGAGAGCCAGGGCTGAGTCCAGTTCGAGGAGCAGCAGAG 25858
Qy 2219 GCCAGGGCTGAGAGCCCACTGCCAGAGTCAAGGCTGTGCCCTGAGGTCGGCAGG 2278
Db 25859 GCCAGGGCTGAGAGCCCACTGCCAGAGTCAAGGCTGTGCCCTGAGGTCGGCAGG 25918
Qy 2279 GCTCCACAGGAGCTCTCCAGAGCCCACTGCCAGGAGCAGTCCGGCAGGTCCTCCCTG 2338
Db 25919 GCTCCACAGGAGCTCTCCAGAGCCCACTGCCAGGAGCAGTCCGGCAGGTCCTCCCTG 25978
Qy 2339 GTGCAGATCCGGGAGCTCTGAGGTGATCGGAGGCGCGGACCAATATCCCTGACAT 2398
Db 25979 GTGCAGATCCGGGAGCTCTGAGGTGATCGGAGGCGCGGACCAATATCCCTGACAT 26038
Qy 2399 TCCGAGGTGGACCCCGCTACTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCA 2458
Db 26039 TCCGAGGTGGACCCCGCTACTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCA 26098
Qy 2459 TTCGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCCGCTCCCCC 2518
Db 26099 TTCGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCCGCTCCCCC 26158
Qy 2519 ATGGCTGAGGAGGAGCTGGCGAGTTCCCGAGAGCCACGTGGCCCTGGCCAGGTGAAG 2578
Db 26159 ATGGCTGAGGAGGAGCTGGCGAGTTCCCGAGAGCCACGTGGCCCTGGCCAGGTGAAG 26218
Qy 2579 GGGCCCCACGAGGCTGGAGATCACAGAGGATCAGAGGATGTGGACGCGCTCTGGCA 2638
Db 26219 GGGCCCCACGAGGCTGGAGATCACAGAGGATCAGAGGATGTGGACGCGCTCTGGCA 26278
Qy 2639 GAGGCTGCGGTGGGAGGAGCGAAGTGGTCTCGCCGCTCAGCAGGCTCTTCCACTTC 2698
Db 26279 GAGGCTGCGGTGGGAGGAGCGAAGTGGTCTCGCCGCTCAGCAGGCTCTTCCACTTC 26338
Qy 2699 CCTGGAGGACCTGCGGCTGGATGAGCTGAGAGCTGGGGCTGCGTGAGAGGTGAAG 2758
Db 26339 CCTGGAGGACCTGCGGCTGGATGAGCTGAGAGCTGGGGCTGCGTGAGAGGTGAAG 26398
Qy 2759 GCCTCCGTGGAGACATCTCCCGATCTGAAAGGCGAGGCTGAGGT 2806
Db 26399 GCCTCCGTGGAGACATCTCCCGATCTGAAAGGCGAGGCTGAGGT 26446

RESULT 13
LOCUS CQ730656 18524 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 16590 from Patent W002068579.
ACCESSION CQ730656
VERSION CQ730656.1 GI:42305092
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kts, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: W0 02068579-A 16590 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1. .18524
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 57.9%; Score 1738.8; DB 6; Length 18524;

Best Local Similarity 82.2%; Pred. No. 8.5e-270;
Matches 2277; Conservative 0; Mismatches 37; Indels 456; Gaps 6;
Qy 257 CTGGACATAAAGCCCTCTAAACATCTGATGTGATCTCTGCCGGGAAGACATTTAAATC 316
Db 14805 CAGCCCTTTAGCCCTCTAAACATCTGATGTGATCTCTGCCGGGAAGACATTTAAATC 14864
Qy 317 TGGACATTTGGCTTTGGCCAGAAACATACCCAGCAGAGCTGCAGTTTCAGCCAGTACGGC 376
Db 14865 TGGACATTTGGCTTTGGCCAGAAACATACCCAGCAGAGCTGCAGTTTCAGCCAGTACGGC 14924
Qy 377 TCCCTGAGTTTCGTCCTCCCGAGATCATCAGCAGAAACCTGTGAGGAGAGCTTCAGC 436
Db 14925 TCCCTGAGTTTCGTCCTCCCGAGATCATCAGCAGAAACCTGTGAGGAGAGCTTCAGC 14984
Qy 437 ATTTTGGGCCCATGGGTGTCTCTACCTCAGCCTGACCTGCTCATCCCCCATTTGCCGGC 496
Db 14985 ATTTT-----GCCTGACCTGCTCATCCCCCATTTGCCGGC 15017
Qy 497 GAGAGTACCGTGCACCCCTCTGAAAGTCTCTGAGGGGCGCGTGTCTATGAGCAGCCCC 556
Db 15018 GAGAGTACCGTGCACCCCTCTGAAAGTCTCTGAGGGGCGCGTGTCTATGAGCAGCCCC 15077
Qy 557 ATGCTGCCACCTCAGCAGAGAGCCAAAGACTTTCATCAAGGCTACGCTGCAGAGGCC 616
Db 15078 ATGCTGCCACCTCAGCAGAGAGCCAAAGACTTTCATCAAGGCTACGCTGCAGAGGCC 15137
Qy 617 CCTCAGG-----CCCGGCTTAGTGGCGGCCAGTGGCTCTTCCACCCCTCG----- 661
Db 15138 CCTCAGGAATCCAGGGCTGGGGCTGCCAGAGGGCTTCCATCATGAGGAGGGGCCCTA 15197
Qy 662 -----TTCTGAAATCCATG 676
Db 15198 TCAGGGCACCATGGTCTCTCCCTGCTGCTGGCTGCCACCTCGTCTCCAGAAATCCATG 15257
Qy 677 CTTGGAGGAGGAGCCCACTTCAACACCAAGCAGACTCAAGTTCCTCTCGCCGCGAAGT 736
Db 15258 CTTGGAGGAGGAGCCCACTTCAACACCAAGCAGACTCAAGTTCCTCTCGCCGCGAAGT 15317
Qy 737 CGCTGGCAGCGTTCCCTGATGAGCTACAAGTCCATCTCTGGTGTATCGCTCCATCCCTGAG 796
Db 15318 CGCTGGCAGCGTTCCCTGATGAGCTACAAGTCCATCTCTGGTGTATCGCTCCATCCCTGAG 15377
Qy 797 CTGCTGGGGGCCCAACCCGACAGCCCTCTCTCGGCTAGCCCGGACCTCTGAGAGGAC 856
Db 15378 CTGCTGGGGGCCCAACCCGACAGCCCTCTCTCGGCTAGCCCGGACCTCTGAGAGGAC 15437
Qy 857 ACTGGTGGCTCTCCAGTTCTCTCTCTGACACAGAGCTCGCCCAATTTGCCCGG 916
Db 15438 ACTGGTGGCTCTCCAGTTCTCTCTCTGACACAGAGCTCGCCCAATTTGCCCGG 15497
Qy 917 GCTAAGTCACTGGCCACCCCTCCCGGCTGACACACTCACCACCTGCTGCACCCCGGGGCTTC 976
Db 15498 GCTAAGTCACTGGCCACCCCTCCCGGCTGACACACTCACCACCTGCTGCACCCCGGGGCTTC 15557
Qy 977 CTGGGGCCCTCGGCGAGCTGCTCTGAGAGAGCCGAGGCACTGAGCGCTCCACCGAGGCC 1036
Db 15558 CTGGGGCCCTCGGCGAGCTGCTCTGAGAGAGCCGAGGCACTGAGCGCTCCACCGAGGCC 15617
Qy 1037 CCAGCTCGGCTGATCTCCGAGGGTCCGGGCGCCAGCCGCGGCTCGGTGCCC 1096
Db 15618 CCAGCTCGGCTGATCTCCGAGGGTCCGGGCGCCAGCCGCGGCTCGGTGCCC 15677
Qy 1097 CGGCAACAGCGTCACTCCGAGCTTCTTACCAACAGCGGGTGTAGAGCCCTGTAGCACGGG 1156
Db 15678 CGGCAACAGCGTCACTCCGAGCTTCTTACCAACAGCGGGTGTAGAGCCCTGTAGCACGGG 15737
Qy 1157 GCCTTGGCCCGGGAGAGAGCGGCAACCGGCGCGGCGGCGGCACTGTCTGAAAGGCGGG 1216
Db 15738 GCCTTGGCCCGGGAGAGAGCGGCAACCGGCGCGGCGGCGGCACTGTCTGAAAGGCGGG 15797
Qy 1217 TACATTTGGGGGGCGCTGCCAGGGCTGGCGAGCCACTGATGAGACCGCGCTGTGAG 1276

COMMENT

On Jul 7, 2000 this sequence version replaced gi:8748947.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0661B12
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2478: contig of 2478 bp in length
* 2479 2578: gap of unknown length
* 2579 6948: contig of 4370 bp in length
* 6949 7048: gap of unknown length
* 7049 9373: contig of 2325 bp in length
* 9374 9474: gap of unknown length
* 9474 13303: contig of 3830 bp in length
* 13304 13404: gap of unknown length
* 13404 18270: contig of 4867 bp in length
* 18271 18370: gap of unknown length
* 18371 25589: contig of 7219 bp in length
* 25590 25689: gap of unknown length
* 25690 34815: contig of 9126 bp in length
* 34816 34915: gap of unknown length
* 34916 45742: contig of 10827 bp in length
* 45743 45842: gap of unknown length
* 45843 62987: contig of 17145 bp in length
* 62988 63088: gap of unknown length
* 63088 87584: contig of 24497 bp in length
* 87585 87684: gap of unknown length
* 87685 114765: contig of 27081 bp in length
* 114766 114865: gap of unknown length
* 114866 142334: contig of 27469 bp in length
* 142335 142435: gap of unknown length
* 142435 174612: contig of 32178 bp in length.

FEATURES

source
1..174612
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clones="RP11-661B12"
1..2478
/note="assembly_name:Contig4"
2579..6948
/note="assembly_name:Contig5"
7049..9373
/note="assembly_name:Contig6"
9474..13303
/note="assembly_name:Contig7"
13404..18270
/note="assembly_name:Contig8"
18371..25589

misc_feature
/note="assembly_name:Contig9"
25690..34815
misc_feature
/note="assembly_name:Contig10"
34916..45742
misc_feature
/note="assembly_name:Contig11"
45843..62987
/note="assembly_name:Contig12
clone_end:SP6
vector_side:right"
misc_feature
/note="assembly_name:Contig13"
63088..87584
misc_feature
/note="assembly_name:Contig14"
87685..114765
misc_feature
/note="assembly_name:Contig15"
114866..142334
misc_feature
/note="assembly_name:Contig16"
142435..174612
ORIGIN
Query Match 50.5%; Score 1517; DB 2; Length 174612;
Best Local Similarity 97.2%; Pred. No. 2e-234;
Matches 1641; Conservative 0; Mismatches 35; Indels 13; Gaps 9;
Qy 1118 CTGTTTACACACAGCGGCTGAGAGCCCTGAGACACGGGGCCCTGGCCCGGGGAGCAGG 1177
Db 25403 CTTTCTTACACACAGCGCTGATGAACCCCT--GCAGGGCGCTCTGGCTCCGGGCACACAG 25346
Qy 1178 CGGACACCGCCCGCGGGCGGCACTGCTGAAGGGGGGTACATTGCGGGGGCGGTGCCA 1237
Db 25345 CGAACCCTGG--GCACAGTCGGCACCTGCTGAAGGGGGGTACATTGCGGGG--GCTGCCA 25289
Qy 1238 GGCTGCGGAGCCACTGATGGAGCACCGGCTGCTGGAGGAGGAGCGCCGACGAGGAG 1297
Db 25288 GGCCTGGCGAGCCACTGATGGAAGACAGCGTCTGGAGGAGGAGCGCCGACGAGGAG 25229
Qy 1298 CAGGCCACCCCTCTGGGCCAAAGCCCTCATTTGAGACTGCCCTCCGGCTGCTGCTCT 1357
Db 25228 CAGGCCACCCCTCTGGGCCAAAGCCCTCATTTGAGACTGCCCTCCGGCTGCTGCTCT 25169
Qy 1358 GGCACCCACTTGGCCCTCTGGCCACAGGCACTCCCTGGAACATGACTCTCGAGACACCCC 1417
Db 25168 GG-ACCCACTTGGCCCTCTGGCCACAGCCACTCCCTGGAACATGACTCTCCACGACCCCC 25110
Qy 1418 CGCCCTCTCGGAGGCTCGGCTGAGGACACAGCACTGCTTCAGCCCTCCCGGGGGG 1477
Db 25109 CG-CCCTCTCGGAGGCTCGGCTGAGGACAGGCACTGCTTCAGCCCTCCCGGGGGG 25053
Qy 1478 GCCCTTATCAGGACATGGGGCACCTTCAGGGCTCCAAGCAGCTTCATCCATCGTGGC 1537
Db 25052 CCCTTATCAGGACATGGGGCA-CCTCAGGGCTCCAAGCAGCTTCATCCATCGTGGC 24994
Qy 1538 CACCCAGGCACTGCTCAGCAGAGAGGCGCATCCCGGACAGCCCTTGGGGGAGCCAGCC 1597
Db 24993 CACCCAGGCACTGCTCAGCAGAGAGGCGCATCCCGGACAGCCCTTGGGGGAGCCAGCC 24934
Qy 1598 CTTTCTGCGCACCCCAAGCAGGAGTCTGCGCCCGCCAGAGGGCTGACGCCCCCAGCA 1657
Db 24933 CTTTCTGCGCACCCCAAGCAGGAGTCTGCGCCCGCCAGAGGGCTGAG--CCACCCAGCA 24876
Qy 1658 GTTGGCCCATGCGCTCTGCTGCTCTCTCCTCAGGATCTTGAAGAGAGCCCTTTAGTA 1717
Db 24875 GTTGTTCATGCGCTCTGCTGCTCTCTCCTCCTCCAGGATCTTGAAGAGAGACCCCTTAGTA 24816
Qy 1718 CCCTCAAGCCCTTCTTGGGACAGCCCGGACAGCCCGCTGCGCCCGCCCAAGAGAGCCCCC 1777
Db 24815 CCCTCAAG--CCCTTCTTGGGACAGCCCGGACAGCCCGCTGCGCCCGCCCAAGAGAGCCCCC 24757
Qy 1778 CCATTGGACTTAAGATGGGGCTGGAGACATCTCTCTCTCTGGGAGGAGCAAAACCCGCG 1837
Db 24756 CCATTGGACTTAAGATGGGGCTGGAGACATCTCTCTCTCTGGGAGGAGCAAAACCCGCG 24697
Qy 1838 CCCTGAGTTCCTCCAGGGGTACGCTCCAGGAGGAGCTCTTCCCAAGTGAAGTCCCTCAGG 1897

Db	24696	CCCTGCAGTTCCCCAGGGTCAGCCTCCACGGGAGCTCTTCCCAAGTGAGCTCCCTCAGG	24637
Qy	1998	GTGGGCTCTCCCAAGTGGGACACAGAGCCTGGCCCCCTCCCTGATGCGGAGGGCTGGACC	1957
Db	24636	GTGGGCTCTCCCAAGTGGGACACAGAGCCTGGCCCCCTCCCTGATGCGGAGGGCTGGACC	24577
Qy	1958	CAGGAGGCTGAGGATCTGTCCGACTCCACACCCACCTTGACGCGCCTCAGGACAGGCG	2017
Db	24576	CAGGAGGCTGAGGATCTGTCCGACTCCACACCCACCTTGACGCGCCTCAGGACAGGATG	24517
Qy	2018	ACCATGGCCAAAGTTCTCCCTGGGTGGTGGCGGGGCTACGACGGCGTGGCTGGCTATGGC	2077
Db	24516	ACCATGGCCAAAGTTCTCCCTGGGTGGTGGCGGGCTACGACGGCGTGGCTGGCTATGGC	24457
Qy	2078	ACCTTTGGCTTTGGTGGAGATGACAGGGGCAATGCTGGGGACAGGGCCCATGTGGGCCAAG	2137
Db	24456	ACCTTTGGCTTTGGTGGAGATGACAGGGGCAATGCTGGGGACAGGGCCCATGTGGGCCAAG	24397
Qy	2138	ATAGCTGGGCTGTGTCCAGTCCGAGGAGGAGCAGGAGGAGGCGCAGGGCTGAGTCC	2197
Db	24396	ATAGCTGGGCTGTGTCCAGTCCGAGGAGGAGCAGGAGGAGGCGCAGGGCTGAGTCC	24337
Qy	2198	CAGTCGAGGAGCAGCAGAGGCGCCAGAGCCTGAGAGCCCACTGCCCCAGGTCAGTGCAGG	2257
Db	24336	CAGTCGAGGAGCAGCAGAGGCGCCAGAGCCTGAGAGCCCACTGCCCCAGGTCAGTGCAGG	24277
Qy	2258	CCTGTGCTGAGGTCGGGAGGGCTCCCAACAGAGCTCTCAGAGCCCAACCCCATGGGAG	2317
Db	24276	CCTGTGCTGAGGTCGGGAGGGCTCCCAACAGAGCTCTCAGAGCCCAACCCCATGGGAG	24217
Qy	2318	GACATCGGGAGGTCCTCCCTGGTCAGATCCGGGACCTGTGAGGTGATGCGGAGCGGCC	2377
Db	24216	GACATCGGGAGGTCCTCCCTGGTCAGATCCGGGACCTGTGAGGTGATGCGGAGCGGCC	24157
Qy	2378	GACACAAATATCCCTGGACATTTCCGAGGTGACCCCGCTACCTCAACCTCTCAGACCTG	2437
Db	24156	GACACAAATATCCCTGGACATTTCCGAGGTGACCCCGCTACCTCAACCTCTCAGACCTG	24097
Qy	2438	TACGATATCAAGTACCTCCCAATTCGAGTTTATGATCTTTCAGGAAAGTCCCAAGTCCGCT	2497
Db	24096	TACGATATCAAGTACCTCCCAATTCGAGTTTATGATCTTTCAGGAAAGTCCCAAGTCCGCT	24037
Qy	2498	CAGCCAGAGCGCCTCCCTCCCATGGCTGAGGAGGCTGGCGGATTCCTCCGGAGCCCAAG	2557
Db	24036	CAGCCAGAGCGCCTCCCTCCCATGGCTGAGGAGGCTGGCGGATTCCTCCGGAGCCCAAG	23977
Qy	2558	TGGCCCTGGCCAGGTGAACCTGGGCCCCCAGCAGGCTTGAGATCACAGAGGATCAGAG	2617
Db	23976	TGGCCCTGGCCAGGTGAACCTGGGCCCCCAGCAGGCTTGAGATCACAGAGGATCAGAG	23917
Qy	2618	GATGTGACGCGCTGCTGGCAGAGGCTGCCCTGGGACAGGCAAGTGTCTTCGCGG	2677
Db	23916	GATGTGACGCGCTGCTGGCAGAGGCTGCCCTGGGACAGGCAAGTGTCTTCGCGG	23857
Qy	2678	TCAGCAGCCTTCTCCATCCCTGGGAGGCACTGGCGCTGGATGAGCCTGCAGAGCTG	2737
Db	23956	TCAGCAGCCTTCTCCATCCCTGGGAGGCACTGGCGCTGGATGAGCCTGCAGAGCTG	23797
Qy	2738	GGGCTGCTGAGAGAGTGAAGGCTCTCCGTGGAGCACATCTCCCGGATCCTTGAAGGGCAGG	2797
Db	23796	GGGCTGCTGAGAGAGTGAAGGCTCTCCGTGGAGCACATCTCCCGGATCCTTGAAGGGCAGG	23737
Qy	2798	CCGGAAGGT 2806	
Db	23736	CCGGAAGGT 23728	

Search completed: March 21, 2005, 05:35:32
Job time : 13553.8 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:24:10 ; Search time 4515.62 Seconds
(without alignments)
10741.310 Million cell updates/sec

Title: US-10-077-130-6_COPY_2000_3000
Perfect score: 1001
Sequence: 1 cgcctaccatccggaggtg.....ctgagcaggggccaatgcc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ats.*
12: gb_sv.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1001	100.0	20435	9 HSAJ2535
2	830.8	89.0	2768	6 ARS42178
3	531.2	53.1	18524	6 CQ730656
4	284.8	28.5	24545	9 HSA314898
5	284.8	28.5	135964	9 AL353593
6	229.6	22.9	19475	9 HSA314900
7	225.4	22.5	260998	2 AC099089
8	224	22.4	2488	6 BD160445
9	224	22.4	2488	6 AX883644
10	224	22.4	2488	9 AC024186
11	224	22.4	174612	2 AC023889
12	221.6	22.1	3956	6 C0851220
13	220.6	22.1	3956	9 AK128447
14	220.2	22.0	244553	2 AC098133
15	220	22.0	6400	9 HSA314903
16	218.8	21.9	211829	10 AL645854
17	217.6	21.7	10091	9 HSA314901
18	213.8	21.4	44358	2 AC142478
19	212.2	21.2	164766	2 AC026657

C	20	186.6	18.6	250087	2	AC096931	AC096931 Rattus no
C	21	151.8	15.2	526	11	BV077840	BV077840 S208P6660
	22	112.8	11.3	5783	6	CQ715191	CQ715191 Sequence
	23	98.4	9.8	724	6	BD150413	BD150413 Primer fo
	24	98.4	9.8	724	6	AX870351	AX870351 Sequence
	25	98.4	9.8	2184	6	BD160248	BD160248 Primer fo
	26	98.4	9.8	2184	6	AX883309	AX883309 Sequence
	27	98.4	9.8	2184	9	AK023854	AK023854 Homo sapi
	28	93	9.3	5314	10	AK122340	AK122340 Mus muscu
	29	92.4	9.2	5382	9	AB014557	AB014557 Homo sapi
C	30	91.2	9.1	168144	2	CR792456	CR792456 Danio rer
	31	90.8	9.1	5378	6	CQ875354	CQ875354 Sequence
	32	90	9.0	1704	9	AF035292	AF035292 Homo sapi
	33	88.4	8.8	2025	9	AK025946	AK025946 Homo sapi
	34	88.4	8.8	3316	9	BC007201	BC007201 Homo sapi
	35	87.8	8.8	2761	9	BC043916	BC043916 Homo sapi
	36	86.2	8.6	2534	6	AX834822	AX834822 Sequence
	37	86.2	8.6	2534	9	AK097489	AK097489 Homo sapi
	38	84.2	8.4	3015	10	BC056205	BC056205 Mus muscu
	39	81.4	8.1	933	9	BC061909	BC061909 Homo sapi
	40	75	7.5	2417	10	BC051643	BC051643 Mus muscu
C	41	70	7.0	200465	2	AC148786	AC148786 Otolemur
C	42	70	7.0	215101	2	AC148847	AC148847 Otolemur
	43	69.2	6.9	125020	9	AF429315	AF429315 Homo sapi
C	44	66.8	6.7	125020	9	AF429315	AF429315 Homo sapi
C	45	66.6	6.7	193064	2	AC149860	AC149860 Papio anu

ALIGNMENTS

RESULT 1	HSJAJ2535	Homo sapiens mRNA for obscurin (OBSCN gene).	20435 bp	mRNA	linear	PRI 14-SEP-2001
LOCUS	HSJAJ2535					
DEFINITION	Homo sapiens mRNA for obscurin (OBSCN gene).					
ACCESSION	AJ002535					
VERSION	AJ002535.1	GI:15026973				
KEYWORDS	OBSCN gene; obscurin.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	1	Young, P., Ehler, E. and Gautel, M.				
AUTHORS	1	Young, P., Ehler, E. and Gautel, M.				
TITLE	1	Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly				
JOURNAL	1	J. Cell Biol. 154 (1), 123-136 (2001)				
MEDLINE	21342081					
PUBMED	11448995					
REFERENCE	2	(bases 1 to 20435)				
AUTHORS	Gautel, M.S.					
TITLE	Direct Submission					
JOURNAL	Submitted (31-OCT-1997)	Gautel M.S., Structural Biology Division, European Molecular Biology Laboratory, Meierhofstr. 1, Heidelberg, 69117, GERMANY				
REMARK	Revised by author 20-JUL-2001					
FEATURES	Location/Qualifiers					
source	1..20435					
	/organism="Homo sapiens"					
	/mol_type="mRNA"					
	/db_xref="taxon:9606"					
	/chromosome="1"					
	/map="1q42.1"					
	/cell_type="cardiomyocyte"					
	/tissue_type="cardiac"					
	/dev_stage="adult"					
gene	1..20435					
	/gene="OBSCN"					
CDS	72..19934					
	/gene="OBSCN"					
	/function="titin binding"					
	/codon_start=1					
	/evidence=experimental					

/product="obscurin"
/protein_id="CAC4768.1"
/db_xref="GI:15026974"
/db_xref="GOA:O96AA2"
/db_xref="UniProt/TREMBL:O96AA2"
/translation="MDQPFSGAPRLFRPKAFVVSQKDATLSQIVGNTPPQVSWE
KQQPVTAGPRLAQDGLYKLTLLDALGDSQGYVCRARNALGEAFPAVLQGLVD
AACQAPFHLRLPTISIRREGSEATFRGQSVSPAPSWSKDGRRLGDEGPRVRV
BELGASALRTRAAPRPGDGYEVRANPLGAASAAALVVDSDAATASRPGTSTA
LHLLQRRREARMEAGCAPSPSTRTCTTEGKHARLSCVVTGEKPEPTVMKKDQ
LVTEGRHVVEDAKENVLKILCKQSDRLGYLTASNLGQYSSVLVVRPAVP
PKRLDLVREKASEATFLCEVPQPSSTEANMKETRLWASAKYGBIEGKERLTVR
NYSADDDAYICETPEGSRTVAELAVQGNLRLKPKRTAVAGTMAPCEBLAVPVG
VHMLRQBEVAGRGVATISAGETRHLTISQCLEEDQVAFMAGDCQTSFRFCVSAP
BKPLQPPVPTVKARMESSVILSWSPPHGPERPVTIDGYLVKKKLTGYTICH
EWAATPELTVADEAGNFRVSAINSGQSPYLEPPTGTVHLAKLAVRPLKAVOA
VEGGRVTSVDLTVASAGEMFLDGOALKASSVRIHCDTRBHTLTIREVPSLHGAQL
KFAVGISSIRMEVRAAPGLTANKPPAPAAAREVLARLHESQAALLAELSDQAAVTWL
KDRTLSPGRTVEQASAGRVLVLRDVARDDAGLYECVSRGRIAYQLSVOGLARL
KDMWAGSCDVAAGAPQFECETSEAHVHVHMDGMBELGHSGERFLQEDYGTGRLV
AATVTRQDEGYSICRVGSDSDVFLRVSEPKVFAKEQLARLKLQABAGASATLSCEV
AAQOTETVWYKDGKLLSSSKVMEATGCTRLRVVQOAGADAGEYSCAAGORLSFH
LDVKEPKVFAKDVHAHSEVQAEAGANATLSCEVAQAQAEVWYKDGKLLSSSKLVHV
EAKCRRELIVQQAQKLTAGDYSCEARGQVSFRLHITPEPMWFAKQCSVHNEVQABA
CASAMLSCEVAQOTEVWYKDGKLLSSSKVMEVKGCTRLRVLPQAGADAGEYSC
BAGGQVSFHLHIITPEKGVFAKEQSVHNEVQAEAGTAMLSCEVAQOTEVWYKDGK
KLSSSKVRMEVKGCTRLRVLPQAGADAGEYSCAAGQVSFOLHITPEKAVPAKEQ
LVHNEVTEAGASATLSCEVAQOTEVWYKDGKLLSSSKVRTEAAGCMROLVVQQA
QOADAGEVTCAGGQRLSFHLDVSEPKVFAKEQLARHVKQAEAGATLSCEVAQ
TEVWYKDGKLLSSSKVRMEVAGCTRLRVVQQAQADTGEYSCAAGQRLSFSLDVA
EPKVVFAKEQPVHREVAQASATLSCEVAQOTEVWYKDGKLLSSSKVRMEVAGT
TLSCEVAQOTEVWYKDGKLLSSSKVRMEVAGCTRLRVVQAEAGADAGEYSCAGD
QRLSFHLHVAQTEVWYKDGKLLSSSKVRMEVAGCTRLRVVQAEAGADAGEYSCAGD
SKVRVAVGCTRLRVVQQAQADAGEYSCAAGQRLSFRLHVAELSPQISERCPRE
PLVKEHEDIILITATLTPSAATVWYKDGKLLSSSKVRMEVAGCTRLRVLPQAGAD
DSALYSCRVAAGQDPVQVEEAAKFCRLLEPFSKMSLEGMAELRLRLMPADAGYV
CMGTPRVRKGFQWAGPVRVSLVGLRAEDAGEYSCESRDHTSALQSVSPRVVYK
FMSGLSTVVAEGEAGTFCQVSPDVAVMVRDGLALQPSSEKFAISQSGASHSLTIS
DLVLEDATQITVEAEGSSAAELARREAPVLFKKLEPQTVBERSSVTLVELTRPWP
ELRWTRNATALPCGNVEIIHAEGARHLVLHNTGVFADRGFCGCTPDQKTLVEM
ROVLRLGLOAVEARQATMEVQLSHADVDGSWTRDGLRFQGGPTCHLAVRGMHT
LTLGLRPEDSGLMVFKAQGVTSARLVTVLPSYRCPQDQVVTTEKKEVLTCELS
RPNVDRLWKDGLBELKATMAAAGACSLTLYIFRCFADQGVQVQADHAQSSASV
KQVQRTLYLRYRLVADAGEIQFVAENAESRAQLRVKELPVLTVLRPLRDLAMEKHR
GVLEQVSRAAEWFGQSEQLQPGPKYELVSDGLYKRLIISDVHAEDBDTVCDAQ
DNKTSQAFVVEQSIITVIRGLQDVTVMPEAPWPECETSIPSVRPPKWLIGKTVLQAG
VNGLEQEGTVHRLMRLTCTMTGTVHFTVGKRSARLVVSDIPVWLRPLPQKTVY
RELOSIVLSCDFRPAKAVQYKDDTFLPSFKFMSLEGMAELRLRLMPADAGYV
RCQGAHSSTEVTVAREVTVTQLODAEATEGWASFSCELSHEDVEVWELNGMP
LYNDSFHEI SHKGRHTLVLSIQRADAGIVRASLSKVSTARLEVRKVPVFLKALD
DLSAERETLALQCEVSDPEAHVVRKDGVLQPSDKYDFLHTAGTRGLVVDHVSPE
AGLYTCHVGSEETRARVRVHDLHVGITKRLKTWEVLEGESCPFCVLVSHESADPAMW
TVGGKTVSSSRFOATQGRKYILVREAPSDAGEVVFVSRGLTSKASLIVRPA
IIKPLEDQWVAPGEDVRLCELRAQTPVHMLKDKRAIKRSQKTDVVEGTMAMLVIR
GASLDAGEYTCVEASKSTASLHVEEKANCFTEELNLQVEEKQDIAVFTCKTEHPAA
TVWTKGLLELRAGSGRQSQEGTLRLTISALEKADSDDYTCIDIGQAQRAQLLVQ
RVVHIIDLEEDVQEGSSATFRICRISPANVEPVHWPDLTPLHNAELNIDIAQPGY
HVLTLROLALKDGTIYFEAGDQASAAALVTEKPSVFSRELTDATITEGEDTLVCE
TSTCDIPMCWTQKGLRGSARCOLSHEGHAQLLITGATLQDSGRYKCBAGACSS
IVRVHARPVRQBAKLDLEVEGAATLRCVLSVAPRWKCYGNVIRPDGKYSLRQ
EGAMLELVRLNRPQDSGRYSQFGDQTTSATLTVTALPAQFIOGKLNKEATEGAT
LRCELSATPVRKGSSETLRDGRDCLRGDQAMCELOIRGLAMVDAEYSCVGBER
TSASATLRPMFAHPIGRLHQESTEGATLRCELSKAAAPVWRKGRSRLRDSHSL
RODQAVCELOICGLAVADAGEYSCVCEERTSATLVKALPAKFTGLRNEBVEGAT
AMLWCELSKAPVWRKGPENLRDGRVILRQBSTCELOICGLMAWAGEYLVCGQ
ERLTSATLIRALPARFIVDYNQBEARGATAVLQCELSNAPVWRKGSSETLRDGRY
SLRQDGTKELOIRGLAWADTGEYSCVCGQERTSAMLTVRALPIKFTEGRLNEBTEG
ATAVLRCELSKMAPVWRKGHETLRDGRHSRLQDQARCELOIRGLVAEDAGEYLCMC
GKERTSAMLTVRAMPKFIIEGLRNEATEGDTATLWCELSKAAAPVWRKGHETLRDQ
RHSLRQDSRCELSQIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIVDYNQBAR
EGATAVLQCELSKAAAPVWRKGSSETLRDGRYSRLRQDGTKELOIRGLSVADTGEYSC
VCCQERTSATLTVRAQPVPRPQLSQAEGSATLQCELSSEPTATVWMSGLQLQ

ORIGIN

Query Match	100.0%	Score 1001;	DB 9;	Length 20435;
Best Local Similarity	100.0%;	Pred. No. 1.1e-158;		
Matches 1001;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CGCTACCATCGGGAGGTGCCCGCAGCTTGCACGGGGCGAGCTGAAGTTCGTGCCCA	60	
Db	2071	CGCTACCATCGGGAGGTGCCCGCAGCTTGCACGGGGCGAGCTGAAGTTCGTGCCCA	2130	
Qy	61	ACGGCAITCAGACACATCCGATCGAGTCCGGCGGCGCCCGAGGCTGACTGCCAACA	120	
Db	2131	ACGGCAITCAGACACATCCGATCGAGTCCGGCGGCGCCCGAGGCTGACTGCCAACA	2190	
Qy	121	AGCCGACCGCAGCTGCCCGGGAGGTGCTGGCTGCGCTGCAACGAGGAGCGCAGCTGC	180	
Db	2191	AGCCGACCGCAGCTGCCCGGGAGGTGCTGGCTGCGCTGCAACGAGGAGCGCAGCTGC	2250	
Qy	181	TGGCTAGCTGTGATCAGGCTGCGGCTGTGACGTGGCTGAAGGATGGTCGACACATGT	240	
Db	2251	TGGCTAGCTGTGATCAGGCTGCGGCTGTGACGTGGCTGAAGGATGGTCGACACATGT	2310	
Qy	241	CCCCAGCCCCCAAGTATGAGGTGCAGGCATCGCGGGCGGGGGTGTCTCTTGTGCGAG	300	
Db	2311	CCCCAGCCCCCAAGTATGAGGTGCAGGCATCGCGGGCGGGGGTGTCTCTTGTGCGAG	2370	

ANGRREPLRQCTAELVLQDLQREDTGEYTCCTCGSQATSATLTVTAAAPVRLRLQHO
EVDGGTGHLLCCSLSRASVEMRKSGLSQLPFCACKYQMVODAAAYELVVRVQEDAG
DTCDTCTHOSMASLSVRVPRPKFKTFLQSLQEOFTGDIARLCCOLSABESAVQWLK
EGVELHAGPKYEMRSQATRELLIHQLEAKDTGEYACVTGGQKTAASLRVTEPEVTIV
RGLVDATVDEDEVEFSCEVRAGATQVCMQGLPLQSLQESNEVEVAVRDGRHRLRLK
FTQPEDAGTVSFHLGNHSAQLTVRAPVETIIEPLQDLQSEGQDASFCQRLSRASG
QEARWLVGVLQANEMNDITVEQGTLLHLLTKHVTLEDAQTVSFHVTGTCSSAQGLKV
TAKNTVVRLGLENVALEBGBALFCOLSQPEVAHTWLLDDQVPTSENARVPEFNG
LRLHLLKVLNRPQDSCHVTLFAGMDWTSALTVRGWLELLEPLKNAARVAGAARFT
CTLSEALVPEASWYINGAAVQPDSDWTVTADGSHQALLRLSQPHHAGEVTPACRD
AVASARUTVLGLPDPPEDAEVVHSSHTVLSWAAPMSDGGGLCGVVRKVEGATGQ
WRICHELVPQPECVQDGLAFGETVFRVAAVFPVGAEPVHLPTQVRLAEPKPKVPVQ
SPASERQVAAAGEDVLSLEVAEAGEVIWHKMERIQPGFERFEVQGGQMLVIKG
FTADQGEYHCLGAQGSICPAATFOVALPSDEAPQSPSLPEAAQEGDLHLLWEA
LAKRRMREPTLDSISELPEEDGRSQRLPQEAEEVAPDLSEGYSTADELATGDAI
SHTSDDSEAGTSPSLVYLKKAGRPCTSLASKVGAAPAASVKPQQQBPFLAARPP
LGLSLKDLGDPDSMDKAAVLIQAAGFKYKVRKEMKQEQGPMFSHTFGDTAQVGDALR
LECVAASDVARWLKQVLELTDGRHHHLDQLDGTCSLLIAGLDRADAGCTQVRS
NKDGQVTHSACVVVSGSEASESSSGELDDAFRRAARLRLHRLPRTKSPAESDEELF
LSADEGPAEPPEADWQTYREDSHFCIRPEALTTEARQAVTRFOEMFATLIGIVEIKL
VEGPRRVEVICSETPAVVPPEPLPSLLTSDAAPVFLTELQNEQVQDGYPVSFDCV
VTGQPMESVVRWKDKLLEDDHVMINEDQGGHOLIITAVPADMGVYRCLANSWG
YSSTKAELRVDLTSTDYDTHAADATESSYFSAQYLSRREQESTDIBGQLPVVE
ELRDLQVAPQTRLAKTQKVKGYPAPLVWFQDQPLTASAHIRMTDKTLHLTEIIS
VTRDESGQYAAYSNAGRAYSSARLLVRGDEPEEKPAASDVHQLVPPRMLRFTPK
VGRHOCTYTICIASNAAGALCSALHVSGLPKVERQEKVKEALISFLQGTQALISA
QGPOTASFADLQGRKEEPLAAKEALGHLSAEGVTEEFELKLTSTIEMVSAKITQA
KLQVPGSDSDSKTSPASPHGRSPSSIQBSSSESDGDAGEIFDIYVVVADLYL
PLGAQDAILTRREGQVVLDDAAHPLRVLRTRTKTSSPSRQSGVPAYLDRKLSP
EWGAABEPFPGAEVSEDYKARLSVIOELLSQAFVEELOFQSHHQLHLERCPH
VPTAVAGQKAVIFRNVRDIGHFSSFLQELQCDTDDVMVCFIKQAAEPQYLEFV
ELLRNARVQNCALLQEAQVAVSALPQRAENKLVSLMENVYPTGLQALGEPHQHF
IWEAGAPAMPKGNHNVLPFRNHLVLCRPRDRTDVTVSFVRNMKLSIDLND
QVEDRAFEVQERDSVRKYLQARTAIKSSWVKTCIGIQORLALPVRPPDPEE
ELADTAELEGVTKLACRTVTGPKVIVSWYKDGKAVQVDPHHILIIDPDGSCALILDS
LTVGQYKQMFCAASANCSTLKLIVPVRPNKVLASPFVEBGEAQDTCTIEGA
PYOIRWYKDGALLTNGKFTULSEPSGLLVIVIRAAKEDLGLVECELNRILGSAR
ASALRISQPMLOAQOCHREOLVAAVTEQETKVPKVTVIIIEETITTVVKSQRQR
RSPSKSPSPSCSASPLRGLLADPLILPGAQPPRPAEPFGKQVVPVTLPGATP
EASHPALPGLSGQPKWVEVEETIEVRKMKQGVSPTEVESRSSGHLFTLPGATP
GDPNNSNNKLLQAQWAQGTAMVVRVPLFVRDARSGVDMAASGMGSLSEEGTM
EAGBEEGDGDAFTVEESQDTHSLGDRDPKILTHNGRMLTLADLSDYVPEGEETHC
GGPGCAPDPPCEVSIORIEGEPVTVLLSVGHALCPGRLGFLFRPEPPGASPP
GPQVRSLEGTSPILLREAPRVGSAFWTQSFCTRIIRSADSGSSFTTELSTQTQVNF
TVGETVLIHCPRDDEAAQp"

Qy 301 ATGTGGCCCGGACGATCAGGCTCTACGAGTGGTTCAGCGGGGGCCGATCGGCT 360
Db 2371 ATGTGGCCCGGACGATCAGGCTCTACGAGTGGTTCAGCGGGGGCCGATCGGCT 2430
Qy 361 ACCAGCTCTCGTGCAGGCTCGCGGCTTTTCGCAAGGACATGCGGGGCGAGCTGTG 420
Db 2431 ACCAGCTCTCGTGCAGGCTCGCGGCTTTTCGCAAGGACATGCGGGGCGAGCTGTG 2490
Qy 421 TGGATGCGGTGGTGGGGCCCGGGCCAGTTTGTAGTGTGAGACCTCCGAAAGCCACGTC 480
Db 2491 TGGATGCGGTGGTGGGGCCCGGGCCAGTTTGTAGTGTGAGACCTCCGAAAGCCACGTC 2550
Qy 481 ACGTGCACTGTACAGGATGCACTGAGCTGGGCACTCCGCTGAGCGCTTTTTCAGG 540
Db 2551 ACGTGCACTGTACAGGATGCACTGAGCTGGGCACTCCGCTGAGCGCTTTTTCAGG 2610
Qy 541 AGGATGTGGGACCGGCGCCGCTGTGTGGGAGCCACAGTCACAGGCGAGGATGAAGGCA 600
Db 2611 AGGATGTGGGACCGGCGCCGCTGTGTGGGAGCCACAGTCACAGGCGAGGATGAAGGCA 2670
Qy 601 CCTACTCTCGGTGGGCGAGGACTCTGTGGACTTCCGGCTCCGGCTCTCTGAGGCCA 660
Db 2671 CCTACTCTCGGTGGGCGAGGACTCTGTGGACTTCCGGCTCCGGCTCTCTGAGGCCA 2730
Qy 661 AGGTGGTGTGTTGCTAAGGAGCAGCTGGCCACGAGGAGCTCGAGGAGCGAGGAGGCA 720
Db 2731 AGGTGGTGTGTTGCTAAGGAGCAGCTGGCCACGAGGAGCTCGAGGAGCGAGGAGGCA 2790
Qy 721 GTGCCACACTGAGCTGCGAGGTGGCCAGGCGCCAGAGGTGACGTGGTACAGGATG 780
Db 2791 GTGCCACACTGAGCTGCGAGGTGGCCAGGCGCCAGAGGTGACGTGGTACAGGATG 2850
Qy 781 GGAAGAAGCTGAGCTCCAGCTCGAAGTGTGATGGAGCCACAGGTCACGCGCAGGC 840
Db 2851 GGAAGAAGCTGAGCTCCAGCTCGAAGTGTGATGGAGCCACAGGTCACGCGCAGGC 2910
Qy 841 TGGTGTGACAGGCGAGCGGATGCGGGAGTATAGCTGCGAGGCTGGGGCC 900
Db 2911 TGGTGTGACAGGCGAGCGGATGCGGGAGTATAGCTGCGAGGCTGGGGCC 2970
Qy 901 AGCGGCTCTCCATCTCGATGTCAAAGAGCCCAAGGTGTGTTTCCCAAGGACCAAGG 960
Db 2971 AGCGGCTCTCCATCTCGATGTCAAAGAGCCCAAGGTGTGTTTCCCAAGGACCAAGG 3030
Qy 961 TGGCACACAGTGAAGTGAAGCTGAGGCGAGGCGGCAATGCC 1001
Db 3031 TGGCACACAGTGAAGTGAAGCTGAGGCGAGGCGGCAATGCC 3071

RESULT 2
AR542178 2768 bp DNA linear PAT 08-OCT-2004
LOCUS Sequence 426 from patent US 6743619.
DEFINITION
ACCESSION AR542178
VERSION AR542178.1 GI:53934258
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE
1 (bases 1 to 2768)
AUTHORS Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F.,
Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.J., Wehrman, T., Wang, J.-R.,
Wang, D. and Drmanac, R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 6743619-A 426 01-JUN-2004;
FEATURES Location/Qualifiers
source 1..2768
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 83.0%; Score 830.8; DB 6; Length 2768;
Best Local Similarity 95.3%; Pred. No. 5.6e-130;

Matches 856; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
Qy 1 CGCTCACATCCGGAGAGTGCCTCCGACAGCTGCAAGGCGCAGCTGAAGTTCTGTGGCA 60
Db 1861 CGCTCACATCCGGAGAGTGCCTCCGACAGCTGCAAGGCGCAGCTGAAGTTCTGTGGCA 1920
Qy 61 ACGCATTGAGAGCAGATCCGATGAGGTCCGGGGGGCCCGGCTGACTGCAACA 120
Db 1921 ACGCATTGAGAGCAGATCCGATGAGGTCCGGGGGGCCCGGCTGACTGCAACA 1980
Qy 121 AGCCGCGAGCGCAGCTGCCGGAGGTGCTGGCTCGGCTGACAGAGAGCGCAGCTGC 180
Db 1981 AGCCGCGAGCGCAGCTGCCGGAGGTGCTGGCTGACAGAGAGCGCAGCTGC 2040
Qy 181 TGGCTGAGCTGTACAGTACAGCTCGGCTGCGCTGCGCTGCAAGAGTGTGCACTGT 240
Db 2041 TGGCTGAGCTGTACAGTACAGCTCGGCTGCGCTGCGCTGCAAGAGTGTGCACTGT 2100
Qy 241 CCCAGAGCCCAAGTATGAGGTGAGGATCGGCGGGGGCGGCTGCTTGTGGAG 300
Db 2101 CCCAGAGCCCAAGTATGAGGTGAGGATCGGCGGGGGCGGCTGCTTGTGGAG 2160
Qy 301 ATGTGGCCCGGAGCGATGCAAGGCTCTACGAGTGTGCTCAGCGGGGGCGGCTATCGCT 360
Db 2161 ATGTGGCCCGGAGCGATGCAAGGCTCTACGAGTGTGCTCAGCGGGGGCGGCTATCGCT 2220
Qy 361 ACCAGCTCTCGTGCAGGCTCGCGGCTTTTGTGCAAGGACATGCGGGGCGAGCTGTG 420
Db 2221 ACCAGCTCTCGTGCAGGCTCGCGGCTTTTGTGCAAGGACATGCGGGGCGAGCTGTG 2280
Qy 421 TGGATGCGGTGGTGGGGGGCGGCGGCTTGTGAGTGTGAGACCTCCGAAAGCCACGTC 480
Db 2281 TGGATGCGGTGGTGGGGGGCGGCGGCTTGTGAGTGTGAGACCTCCGAAAGCCACGTC 2340
Qy 481 ACGTGCACTGTGTAAGGATGAGTGTGAGTGTGGGCGCTCCGCTGAGCGCTTTCTTGCAGG 540
Db 2341 ACGTGCACTGTGTAAGGATGAGTGTGAGTGTGGGCGCTCCGCTGAGCGCTTTCTTGCAGG 2400
Qy 541 AGGATGTGGGACCGCGGACCGGCTGTGGGAGCCACAGTCACAGGCGAGGATGAAGGCA 600
Db 2401 AGGATGTGGGACCGCGGACCGGCTGTGGGAGCCACAGTCACAGGCGAGGATGAAGGCA 2460
Qy 601 CCTACTCTCGCGGTGGGGGAGGACTGTGTGAGTGTCCGGCTCCGGCTCTCTGAGGCCA 660
Db 2461 CCTACTCTCGCGGTGGGGGAGGACTGTGTGAGTGTCCGGCTCCGGCTCTCTGAGGCCA 2520
Qy 661 AGGTGGTGTGTTGCTAAGGAGCAGCTGGCCACGAGGAGCTCAGGCGAGGCGAGGCGCA 720
Db 2521 AGGTGGTGTGTTGCTAAGGAGCAGCTGGCCACGAGGAGCTCAGGCGAGGCGAGGCGCA 2580
Qy 721 GTGCCACACTGAGCTGCGAGGTGGCCAGGCGCCAGAGCGAGGTGACGTGGTACAGGATG 780
Db 2581 GCGCCAGCTGAGCTGTGAGGTGGCCAGGACAGATGGAGGTGACGTGGTACAGGAGCG 2640
Qy 781 GGAAGAAGCTGAGCTCCAGCTCGAAGTGTGATGGAGGCGCAGGCTGCAAGCGCAGGC 840
Db 2641 GGAAGAAGCTGAGCTCCAGCTCGAAGTGTGATGGAGGCGCTGTGGGCTGTATGCGGAGGC 2700
Qy 841 TGGTGTGTCAGCAGGCGAGGCGAGGATGCGGGGAGTATAGCTGCGAGGCTGGGGG 898
Db 2701 TGGTGTGTCAGCAGGCGAGGCGAGGATGCGGGGAGTATAGCTGCGAGGCTGGGGG 2758

RESULT 3
CO730656 18524 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 16590 from Patent WO02068579.
DEFINITION
ACCESSION CO730656
VERSION CO730656.1 GI:42305092
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 16590 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1. 18524
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 53.1%; Score 531.2; DB 6; Length 18524;
Best Local Similarity 90.7%; Pred. No. 9.1e-80;
Matches 566; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 378 GGCCTCGCGCGCTTTCTGCACAAAGACATGCGCGCAGCTGTGTGGATGCCGTGGCTGGG 437
Db |||||
309 GGCCTCGCGCGCTTTCTGCACAAAGACATGCGCGCAGCTGTGTGGATGCCGTGGCTGGG 368
QY 438 GCGCCGCGCGCAGTTTGTAGTGTGAGACCTCCGAAAGCCACGTCACGTCGACTGTGTACAAG 497
Db |||||
369 GCGCCGCGCGCAGTTTGTAGTGTGAGACCTCCGAAAGCCACGTCACGTCGACTGTGTACAAG 428
QY 498 GATGCGATGGAGCTGGCGCACTCCCGTGTGAGCGCTTCTTTCAGAGGAGATGTGGGACCGGG 557
Db |||||
429 GATGCGATGGAGCTGGCGCACTCCCGTGTGAGCGCTTCTTTCAGAGGAGATGTGGGACCGGG 488
QY 558 CACCGGCTGGTGGCAGCCACAGTCACAGGAGGATGAAGCAGCTACTCTCGCGCGGTG 617
Db |||||
489 CACCGGCTGGTGGCAGCCACAGTCACAGGAGGATGAAGCAGCTACTCTCGCGCGGTG 548
QY 618 GCGCAGGACTCTGTGGACTTCGGCTCCGCTCTCTGAGCCCAAGGTGGTGTTCGTAAG 677
Db |||||
549 GCGCAGGACTCTGTGGACTTCGGCTCCGCTCTCTGAGCCCAAGGTGGTGTTCGTAAG 608
QY 678 GAGCAGCTGGCAGCAGGAAGCTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 737
Db |||||
609 GAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 668
QY 738 GAGTGGCCAGGCGCCAGACGAGGAGTGCAGTGTGACGAGTGAAGGATGGAAGAGAGTGCCTC 797
Db |||||
669 GAGTGGCCAGGCGCCAGACGAGTGCAGTGTGACGAGTGAAGGATGGAAGAGAGTGCCTC 728
QY 798 AGCTCGAAAGTGTGCATGCGAGGCCACAGGCTGCAGCGCAGGCTGTGTGTCAGCAGGCA 857
Db |||||
729 AGCTCGAAAGTGTGCATGCGAGGCCACAGGCTGTGTGTCAGCAGGCTGTGTGTCAGCAGG 788
QY 858 GCGCAGGCGGATGCGCGGAGTATAGCTGCGAGGCTGGGGGCCACAGCGGCTCTCTCTCCAT 917
Db |||||
789 GCGCAGGCGGATGCGCGGAGTATAGCTGCGAGGCTGGGGGCCACAGCGGCTCTCTCTCCG 848
QY 918 CTGATGTCAAAGAGCCAGGAGTGGTGTGTCGCAAGGACAGGTCGACAGTGAAGGTG 977
Db |||||
849 CTGACGCTTCAGAGCCAGGAGTGGTGTGTCGCAAGGAGGAGTGGCAGCAGGAGGAGTGG 908
QY 978 CAGGCTCAGGCGAGGCGCAATGCC 1001
Db |||||
909 CAGGCGAGGCGAGGAGCAGTGGC 932

RESULT 4
HSA314898
LOCUS HSA314898 24545 bp DNA linear PRI 21-MAY-2002
DEFINITION Homo sapiens partial OBSCN gene for obscurin, exons 5-11.
ACCESSION AJ314898
VERSION AJ314898.1 GI:21104327
KEYWORDS OBSCN gene; obscurin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
Young, P., Ehler, E. and Gautel, M.
Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 24545)
TITLE Direct Submission
Gautel, M.S.
JOURNAL Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie,
Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse
11, Dortmund, 44227, GERMANY
FEATURES
source
1. 24545
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
332..10891
/gene="OBSCN"
join(332..592,948..1256,1720..1986,2578..2859,3170..3445,
5481..5750,10616..10891)
/gene="OBSCN"
join(332..592,948..1256,1720..1986,2578..2859,3170..3445,
5481..5750,10616..10891)
/gene="OBSCN"
/codon_start=3
/product="obscurin"
/protein_id="CAC85746.1"
/db_xref="GI:21104328"
/db_xref="UniProt/TREMBL:O8NH8"
/translation="NLLRLPKTAVRVDGTAMFCVELAVPYGPVHWRNQBREVAGG
RYAISAGERTHTLTISQCCLEVDVGQVAFMAGDCQTSTQFCVSAPRKPPLQPPVDPVVK
ARMSSVILWSFPFPHGERPVTIDGVLVEKKLGTITWIRCHEAEVATPELTADV
EENQFVRVSNLNSFGQSPYLEFPFPGTVHLAPKLAVRTPLKVAQVAGGVTFSVDLTV
ASAGEWFLDQALKASSVYEIHCDRTRTLTIREVPASLHGAQLKFAVANGISSIRME
VRAAPGLTANKPPAAAREVLAHLHEEAQLLAELSDQAAATWLDKDRTLSPGPKVEV
QASAGRVLVLDVARDADAGLVECVSRGRIAVQLSVQGLARFLHKDMAGSCVDVAG
GPAQFECSTSEAHVHVHYKDGMEIGHGGERFIQEDVGTGHRLLVAATVTRQDEGTSC
RVGESVDRLRVSPKAVFAKQFPACREVQAEVSGASATLSCEVAQDQMEVWYKDGK
KLSSSKVHVEAVGCMRRLLVVOQGDAGSGEYSCARGQVSRFLDPEKPVFAKQELA
RRKLQAGASATLSCEVAQAEVTWYDKKKLSSSSKVCMEATGCTRRLVYQQAQ
ADAGEYSCBAGGQRLSFHLDFR"
332..592
/gene="OBSCN"
/number=5
593..947
/gene="OBSCN"
/number=5
948..1256
/gene="OBSCN"
/number=6
1257..1719
/gene="OBSCN"
/number=6
1720..1986
/gene="OBSCN"
/number=7
1987..2577
/gene="OBSCN"
/number=7
2578..2859
/gene="OBSCN"
/number=8
2860..3169
/gene="OBSCN"
/number=8
3170..3445
/gene="OBSCN"
/number=9
3446..5480
/gene="OBSCN"
/number=9

```

exon      5481.. 5750
/gene="OBSCN"
/number=10
5751.. 10615
/gene="OBSCN"
/number=10
10616.. 10891
/gene="OBSCN"
/number=11

ORIGIN
Query Match      28.5%; Score 284.8; DB 9; Length 24545;
Best Local Similarity 99.3%; Pred. No. 2.1e-38;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 92 CCGGCGGCCCCCAGGGCTGACTGCCAACAAAGCCGACGCGCGGAGGTGCT 151
Db 2573 CCCAGCGGCCCCAGGGCTGACTGCCAACAAAGCCGACGCGCGGAGGTGCT 2632

Qy 152 GGCTCGGCTGCACGAGGAGGCGCAGCTGCTGGCTGAGCTGTCAGATCAGGCTCGGCTGT 211
Db 2633 GGCTCGGCTGCACGAGGAGGCGCAGCTGCTGGCTGAGCTGTCAGATCAGGCTCGGCTGT 2692

Qy 212 GACGTGGCTGAAGATGTCGCACACTGTCCCGAGGCCCCAAGTATGAGTGCAGGCATC 271
Db 2693 GACGTGGCTGAAGATGTCGCACACTGTCCCGAGGCCCCAAGTATGAGTGCAGGCATC 2752

Qy 272 GCGCGGCGGCGGGTGTCTTGTTCGAGATGTGGCCGCGGACGATGCAGGCCTCTACGA 331
Db 2753 GGCCGCGGCGGCGGGTGTCTTGTTCGAGATGTGGCCGCGGACGATGCAGGCCTCTACGA 2812

Qy 332 GTGGCTGACGCGCGGCGGCGGCGCTACAGCTCTCCGTGCNAGG 379
Db 2813 GTGGCTGACGCGCGGCGGCGGCGCTACAGCTCTCCGTGCNAGG 2860

RESULT 5
AL353593
LOCUS      AL353593      135964 bp      DNA      linear      PRI 13-FEB-2002
DEFINITION Human DNA sequence from clone RP5-1139B12 on chromosome 1q42.1-43,
complete sequence.
ACCESSION AL353593
VERSION   AL353593.33 GI:18673899
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135964)
AUTHORS   Almeida,J.
TITLE     Direct Submission
JOURNAL   Submitted (13-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
COMMENT   On Feb 14, 2002 this sequence version replaced gi:17977879.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

```

```

was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP5-1139B12 is from the library RPCI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-1139B12.
Location/Qualifiers
1..135964
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.1-43"
/clone="RP5-1139B12"
/clone_lib="RPCI-5"
25928..26054
/misc_feature
/note="Sequence from overlapping clone RP11-520H14
(ALJ59510). Assembly confirmed by restriction digest."

ORIGIN
Query Match      28.5%; Score 284.8; DB 9; Length 135964;
Best Local Similarity 99.3%; Pred. No. 1.7e-38;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 92 CCGGCGGCCCCCAGGGCTGACTGCCAACAAAGCCGACGCGCGGAGGTGCT 151
Db 3743 CCCAGCGGCCCCAGGGCTGACTGCCAACAAAGCCGACGCGCGGAGGTGCT 3802

Qy 152 GGCTCGGCTGCACGAGGAGGCGCAGCTGCTGGCTGAGCTGTCAGATCAGGCTCGGCTGT 211
Db 3803 GGCTCGGCTGCACGAGGAGGCGCAGCTGCTGGCTGAGCTGTCAGATCAGGCTCGGCTGT 3862

Qy 212 GACGTGGCTGAAGATGTCGCACACTGTCCCGAGGCCCCAAGTATGAGTGCAGGCATC 271
Db 3863 GACGTGGCTGAAGATGTCGCACACTGTCCCGAGGCCCCAAGTATGAGTGCAGGCATC 3922

Qy 272 GCGCGGCGGCGGGTGTCTTGTTCGAGATGTGGCCGCGGACGATGCAGGCCTCTACGA 331
Db 3923 GGCCGCGGCGGCGGGTGTCTTGTTCGAGATGTGGCCGCGGACGATGCAGGCCTCTACGA 3982

Qy 332 GTGGCTGACGCGCGGCGGCGGCGCTACAGCTCTCCGTGCNAGG 379
Db 3983 GTGGCTGACGCGCGGCGGCGGCGCTACAGCTCTCCGTGCNAGG 4030

RESULT 6
HSA314900      19475 bp      DNA      linear      PRI 21-MAY-2002
LOCUS      HSA314900      19475 bp      DNA      linear      PRI 21-MAY-2002
DEFINITION Homo sapiens partial OBSCN gene for obscurin, exons all-af6.
ACCESSION   AJ314900
VERSION     AJ314900.1 GI:21104329
KEYWORDS    OBSCN gene; obscurin.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS     Young,P., Ehler,E. and Gautel,M.
TITLE       Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere
assembly
JOURNAL     Unpublished
REFERENCE 2 (bases 1 to 19475)
AUTHORS     Gautel,M.S.
TITLE       Direct Submission
JOURNAL     Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie,
Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse
11, Dortmund, 44227, GERMANY
FEATURES
source      1..19475
/organism="Homo sapiens"

```


Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Hoit, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 260998)
Worley, K.C.

Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 260998)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:22855456. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GGLZ
Center clone name: CH230-154E3
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 21929 bases at least Q40
Consensus quality: 21928 bases at least Q30
Consensus quality: 220845 bases at least Q20
Estimated insert size: 226118; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced.
* By the finished sequence as soon as it is available and the accession number will be preserved.
* The accession number of 260998 bp in length.

* 1 260998: contig of 260998 bp in length.
Location/Qualifiers
1. 260998
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-154E3"
1. 1486
/note="wgs_contig"

misc_feature
1. 1486

ORIGIN
Query Match 22.5%; Score 225.4; DB 2; Length 260998;
Best Local Similarity 85.9%; Pred. No. 1.5e-28; Indels 0; Gaps 0;
Matches 250; Conservative 0; Mismatches 41;

Qy 650 CTCTGAGCCCAAGTGGTGGTTTCTAAGGAGCAGCTGGCAGCAGGAGCTGCAGGCGAGA 709
Db 237533 CCCAGAGCCCAAGTGGTGGTTTCTAAGGAGCAGCTGGCAGCAGGAGCTGCAGGCGAGA 237474

Qy 710 GGCAAGGAGCCAGTGCACACTGAGCTGGAGTGGCCAGGCGCCAGAGCGAGGTGACCTG 769
Db 237473 GGCAAGGAGCCAGTGCACACTGAGCTGGAGTGGCCAGGCGCCAGAGCTGAGGTGTCATG 237414

Qy 770 GTACAAGGATGGGAAGAGTGCAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAGGCTG 829
Db 237413 GTTCAAGGAGCGGAAGAAGCTGAGCTCCAGCTCGAAGTGGCATGGAGGCTCGGGCTG 237354

Qy 830 CACGCGAGGCTGGTGTGTCAGGAGGAGCCAGGCGGATGCCGGGAGTATAGCTGCGA 889
Db 237353 CTCCAGGAGGCTGGTGTGTCAGGAGGAGCGGCGGATGCCGGGAGTATAGCTGCGA 237294

Qy 890 GGCTGGGCGGCGAGGCGGTCTCTTCCATCTGATGTCAAAGAGGCCCAAGGT 940
Db 237293 GGCGCGGAGCAGAGGCTCTCTTCCGCTGGAGCTCACAGGTGAGTTGGT 237243

RESULT 8
BD160445 2488 bp DNA linear PAT 17-JAN-2003
LOCUS BD160445
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD160445
VERSION BD160445.1 GI:27866203
KEYWORDS JP 2002191363-A/15288.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2488)
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 15288 09-JUL-2002;
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/15288
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof PH Key
Location/Qualifiers
FT CDS (697)..(1194).
Location/Qualifiers
1. 2488
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
Query Match 22.4%; Score 224; DB 6; Length 2488;
Best Local Similarity 83.6%; Pred. No. 4.6e-28; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 50;

ORIGIN
Qy 650 CTCTGAGCCCAAGTGGTGGTTTCTAAGGAGCAGCTGGCAGCAGGAGCTGCAGGCGAGA 709
Db 891 CCCAGAGCCCAAGTGGTGGTTTCTAAGGAGCAGCTGGCAGCAGGAGGTGCGAGGCTGA 950

Qy 710 GGCAAGGAGCCAGTGCACACTGAGCTGGAGTGGCCAGGCGCCAGAGCGAGGTGACCTG 769
Db 951 GGCGCGGAGCAGTGCACACTGAGCTGGAGTGGCCAGGCGCCAGAGCTGAGGTGACCTG 1010

```

Qy 770 GTACAAGATGGGAAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGAGAGCCACAGGCTG 829
Db 1011 GTACAAGATGGGAAGAGCTGAGTTCCAGCTCGAAAGTGTGCATGAGAGCCGCTGGCTG 1070

Qy 830 CACGCGCAGGCTGTGTGTCAGCAGCAGGCGCCAGGATGCCGGGAGTATAGCTGCGA 889
Db 1071 CACACGGAGGCTGTGTGTCAGCAGCAGGCGCCAGCAGGCGCGGAGTACAGCTGCGA 1130

Qy 890 GGCTGGGGCCAGCGGCTCTCTCCATCTGGATCTCAAAGAGCCCAAGGTGGTGTTCG 949
Db 1131 GGCAGGGGTCAGCAGCTCTCTCTCGCTGCAGGTGGCAGTCACTGCTTTGGGATGC 1190

Qy 950 CAAG 953
Db 1191 TGAG 1194

RESULT 9
AX883644 AX883644 2488 bp DNA linear PAT 17-DEC-2003
LOCUS Sequence 18549 from Patent EP1074617.
DEFINITION AX883644
ACCESSION AX883644
VERSION AX883644.1 GI:40038545
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 18549 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source 1.2488
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
697..1197
note="unnamed protein product"
/codon_start=1
/protein_id="CA832053.1"
/db_xref="GI:40038546"
/translation="MALHSGWRNDCSAGRWMGFSQLHGQCOAWEGSVPTFSRPSY
PRVFLTWAVSLTSCVPPSPSPKPVFAKEQPAHREVQAEAGASATLSCEVAOATE
VTWKDGKLLSSSKRVEAVGCTFRLVVDQAGQAEAGVSCERAGGQSLFRLQVAGQ
CFDGA"

ORIGIN
Query Match 22.4%; Score 224; DB 6; Length 2488;
Best Local Similarity 83.6%; Pred. No. 4.6e-28;
Matches 254; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 650 CTCAGGCCCAAGTGTGTGTTCAGGAGCAGCTGGCAGCAGGAGCTGCAGGAGCAGA 709
Db 891 CCCAGGCCCAAGTGTGTGTTCAGGAGCAGCAGCAGGAGCTGCAGGAGTGA 950

Qy 710 GGCAGGAGCCAGTCCACACTGAGCTCGAGTGGCGCCAGCCAGCGAGGTGACGTG 769
Db 951 GGCAGGAGCCAGTCCACACTGAGCTCGAGTGGCGCCAGCCAGCGAGGTGACGTG 1010

Qy 770 GTACAAGATGGGAAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGAGAGCCACAGGCTG 829
Db 1011 GTACAAGATGGGAAGAGCTGAGTTCCAGCTCGAAAGTGTGCATGAGAGCCGCTGGCTG 1070

Qy 830 CACGCGCAGGCTGTGTGTCAGCAGCAGGCGCCAGGATGCCGGGAGTATAGCTGCGA 889
Db 1071 CACACGGAGGCTGTGTGTCAGCAGCAGGCGCCAGCAGGCGCGGAGTACAGCTGCGA 1130

Qy 890 GGCTGGGGCCAGCGGCTCTCTTCATCTGGATGTCAAAGAGCCCAAGGTGGTGTTCG 949
Db 1131 GGCAGGGGTCAGCAGCTCTCTCTCGCTGCAGGTGGCAGTCACTGCTTTGGGATGC 1190

```

Qy 950 CAAG 953
Db 1191 TGAG 1194

RESULT 10
AK024186
LOCUS Homo sapiens cDNA FLJ14124 fis, clone MAMWA1002498.
DEFINITION AK024186
ACCESSION AK024186
VERSION AK024186.1 GI:10436503
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Ishibashi, T., Yamashita, H.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Katsuta, N., Sato, K.,
Hirakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hata, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Teraashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, F., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Shimura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)

REFERENCE 2
AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,
Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K.,
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Oshima, A.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2488)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
Location/Qualifiers

COMMENT
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
Location/Qualifiers

FEATURES

```

source      1. 2488
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="MAMMA1002498"
            /tissue_type="Mammary gland"
            /clone_lib="MAMMA1"
            /note="Cloning vector: pME18SFL3"
            697..1197
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="BAB14847.1"
            /db_xref="GI:10436504"
            /translation="MALHSGWRNCDSCAGRWGFSQLHGQCAWEGSVPTFSRPSY
            PRVFLWAVSLTSCVPSPKPVFAKEQPAHREVOAAGASATLSCREVAQATE
            VTWYDKGKLSSSKVRVEAVGCTRRLLVVOAGQAEGYSCAEGGQQLSRLQVAG
            CFGDAE"

CDS
Query Match      22.4%; Score 224; DB 9; Length 2488;
Best Local Similarity 83.6%; Pred. No. 4.6e-28;
Matches 254; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 650 CTCTGAGCCCAAGGTGGTGTTCCTAAGGAGCAGCTGGCAGCAGGAGGTGCAGGCAGA 709
    |||||
Db 891 CCCAGAGCCCAAGGTGGTGTTCCTAAGGAGCAGCTGGCAGCAGGAGGTGCAGGCAGA 950
    |||||

Qy 710 GGCAGGAGCCAGTGCACACTGAGCTGCGAGGTGGCCCGAGCCAGCGAGGTGCACGTG 769
    |||||
Db 951 GGCAGGAGCCAGTGCACACTGAGCTGCGAGGTGGCCCGAGCCAGCGAGGTGCACGTG 1010
    |||||

Qy 770 GTACAGATGCGGAAGAAGCTGAGCTCAGCTCGAAGTGTGCATGAGGAGCCAGGCTG 829
    |||||
Db 1011 GTACAAGATGCGGAAGAAGCTGAGTTCAGCTCGAAGTGGCGTGGAGGCGTGGGGTG 1070
    |||||

Qy 830 CACGCGAGGCTGGTGTGACAGCAGGAGCCAGCGCATCCGGGGAGTATAGCTGCGA 889
    |||||
Db 1071 CACAGGAGGCTGGTGTGACAGCAGGAGCCAGCGCATCCGGGGAGTATAGCTGCGA 1130
    |||||

Qy 890 GGCTGGGGGCGAGCGGCTCTCTTCCTCACTCGATGTCTCAAGAGCCCAAGGTGGTGTTC 949
    |||||
Db 1131 GGCAGGGGCTCAGCAGCTCTCTTCCTCGCTCGCAGTGCAGTGCAGTCTTGGGGATGC 1190
    |||||

Qy 950 CAAG 953
    ||
Db 1191 TGAG 1194

RESULT 11
AC023889 Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
LOCUS AC023889
DEFINITION AC023889.3 GI:8969253
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 174612)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174612)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 7, 2000 this sequence version replaced gi:8748947.
----- Genome Center -----

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0661B12
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2478: contig of 2478 bp in length
* 2479 2578: gap of unknown length
* 2579 6948: contig of 4370 bp in length
* 6949 7048: gap of unknown length
* 7049 9373: contig of 2325 bp in length
* 9374 9473: gap of unknown length
* 9474 13303: contig of 3830 bp in length
* 13304 13403: gap of unknown length
* 13404 18270: contig of 4867 bp in length
* 18271 18370: gap of unknown length
* 18371 25589: contig of 7219 bp in length
* 25590 25689: gap of unknown length
* 25690 34815: contig of 9126 bp in length
* 34816 34915: gap of unknown length
* 34916 45742: contig of 10827 bp in length
* 45743 45842: gap of unknown length
* 45843 62987: contig of 17145 bp in length
* 62988 63087: gap of unknown length
* 63088 87584: contig of 24497 bp in length
* 87585 87684: gap of unknown length
* 87685 114765: contig of 27081 bp in length
* 114766 114865: gap of unknown length
* 114866 142334: contig of 27469 bp in length
* 142335 174612: gap of unknown length
* 174612 174612: contig of 32178 bp in length.
Location/Qualifiers
1..174612
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-661B12"
1..2478
/note="assembly_name:Contig4"
2579..6948
/note="assembly_name:Contig5"
7049..9373
/note="assembly_name:Contig6"
9474..13303
/note="assembly_name:Contig7"
13404..18270
/note="assembly_name:Contig8"
18371..25589
/note="assembly_name:Contig9"
25590..34815
/note="assembly_name:Contig10"

```

CDS

/translation="WSPSPGPGPPHSAWHPPPPPPAPPCCMPAAPPWOTPPWALCWTRR
DAGSEAEQTPQPMCPDSOAGLSIIKQCTVTHRAPPTTRAKGLRHVSGTIIIMSIG
NRMRPLCLGLPGWPGHLSPALPITQGHRO"

ORIGIN

Query Match 22.1%; Score 221.6; DB 9; Length 3956;
Best Local Similarity 86.3%; Pred. No. 1.1e-27;
Matches 245; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 647 CGTCTCTAGAGCCCAAGGTGGTGTGTTGCTTAAGAGAGCGTGCACGACGAGAGCTGCAGGC 706
Db 1135 CATCCAGAGCCCAAGGTGGTGTGTTGCTTAAGAGAGCGCGCATGCAGGAGGTGCAGGC 1076

Qy 707 AGAGCGAGGACCGTGCACACTGAGCTGCGAGGTGCCCCAGGCCACAGCGAGGTGAC 766
Db 1075 GGAGCGGGGGCCAGTGCACGCTGAGCTGTGAGGTGCCCCAGGCCACAGTGGAGGTGAC 1016

Qy 767 GTGTGTACAAGGATGGGAAGAGCTGAGCTCCAGCTCGAAAGTGTGCTAGGAGGCCACAGG 826
Db 1015 ATGTACAGAGCGGGAAGAGCTGAGCTCCAGCTCAAAAGTGCATGGAGGCCAGCGG 956

Qy 827 CTCACGCGCAGGCTGTGTGTCAGAGCGGCGGAGCGGATGCCGGGAGTATAGCTG 886
Db 955 CTACACAGGAGCTGTGTGTCAGAGCGGCGGAGCGGATGCCGGGAGTATAGCTG 896

Qy 887 CGAGCTGGGGCCAGCGGCTCTCTTCATCTGGATGTCAAAG 930
Db 895 TGAGCGGGGGCCCAAGCGGTCTCTCCGCTGCACGTGCAG 852

RESULT 14

AC098133
LOCUS AC098133
DEFINITION Rattus norvegicus clone CH230-162B20, WORKING DRAFT SEQUENCE, 2
unordered pieces.
AC098133
VERSION AC098133.5 GI:30522688
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus

REFERENCE

1 (bases 1 to 244553)
Murny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anylabechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Clelland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Gear,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnarsson,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Haviak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulisedge,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Mitosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,

Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemele,O., Okwundu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfamkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Shen,H.,
Sander,M., Savary,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Soderberg,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Ugmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Woodson,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

REFERENCE
AUTHORS

TITLE
JOURNAL

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 244553)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:23267429.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: CH230-162B20

Center clone name: GGMF

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 221927 bases at least Q40

Consensus quality: 223181 bases at least Q30

Consensus quality: 223817 bases at least Q20

Estimated insert size: 229995; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

		1 242725: contig of 242725 bp in length	
*	242726	242825: gap of unknown length	
		* 242826 244553: contig of 1728 bp in length.	
FEATURES		Location/Qualifiers	
source		1..244553	
		/organism="Rattus norvegicus"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:10116"	
		/clone="CH230-162B20"	
		22770..113965	
misc_feature		/note="clone_boundary	
		clone_end:17	
		site:ECORI	
		end_sequence:BH266944"	
misc_feature		23334..24184	
		/note="clone_boundary	
		clone_end:Sp6	
		site:ECORI	
		end_sequence:BH266946"	
ORIGIN			
Query Match		22.0%; Score 220.2; DB 2; Length 244553;	
Best Local Similarity		86.5%; Pred. No. 1.1e-27;	
Matches 243; Conservative		0; Mismatches 38; Indels 0; Gaps 0;	
Qy	650	CTCTGAGCCCAAGTGGTGTGTTGCTTAAGGAGCAGCTGGCACGAGGAGCTGCAGGCAGA	709
Db	995	CCCAGAGCCCAAGTGGTGTGTTGCCAAGGAGCAGCAGCAGCGAGGTGAAGGCAGA	1054
Qy	710	GGCAGAGCCAGTGCACACTGAGCTGCGAGGTGCGCCAGGCCAGCGAGGTGACGTG	769
Db	1055	GSCCGGGGCCAGTGCACACTGAGCTGCGAGGTGCGCCAGGCCAGCGAGGTGACATG	1114
Qy	770	GTACAAGGATGGGAAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAGGCTG	829
Db	1115	GTTCAAGACCGGAGAGAGCTGAGCTCCAGCTTCAAGTGTGAGTGGAGGCTCGGGCTG	1174
Qy	830	CACGCGCAGGCTGTGTTGTCAGCAGGCGCAGGCGGATCCCGGGAGTATAGTCCGA	889
Db	1175	CTCCAGAGGCTGTGTTGTCAGCAGGCGGCGGCAAGCGGATCTGGGAGTACAGCTCGA	1234
Qy	890	GCGTGGGGCCAGCGGCTCTCTTCCATCTCGATGTCAAG	930
Db	1235	GCGCGGGGACAGAAGGTCTCTTCCGCTCGACGTCAAG	1275
RESULT 15			
HSA314903 6400 bp DNA linear PRI 21-MAY-2002			
LOCUS			
DEFINITION			
Homo sapiens partial OBSCN gene for obscurin, exons C-22.			
ACCESSION			
AJ314903			
VERSION			
AJ314903.1 GI:21104333			
KEYWORDS			
OBSCN gene; obscurin.			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1			
REFERENCE			
AUTHORS			
Young, P., Ehler, E. and Gautel, M.			
TITLE			
Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly			
JOURNAL			
Unpublished			
REFERENCE			
2 (bases 1 to 6400)			
Gautel, M.S.			
Direct Submission			
Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie,			
Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse			
11, Dortmund, 44227, GERMANY			
11, Dortmund, 44227, GERMANY			
Location/Qualifiers			
1..6400			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
FEATURES			
source			

gene	2201.. .6263	/gene="OBSCN"	join(<2201..2468,3519..3794,4017..4295,4425..4691,4819..5088,5998..>6263)	/gene="OBSCN"	join(<2201..2468,3519..3794,4017..4295,4425..4691,4819..5088,5998..>6263)	/gene="OBSCN"	codon_start=1	/product="obscurin"	/protein_id="CAC85750.1"	/db_xref="GI:21104334"	/db_xref="UniProt/TREMBL:Q8NHNS"	/translation="VVFAKQDPVHREVQAEAGTSTMLSCVEAQAQTEVVMYKDGKKLS	SSSKMRVEAVGCTRRLLVVBAGQADAGEYSCGAGQRLSFLHLHVAEPKVPFAKEQAPAC	REVQAEAGASATLSCVEAQAQMEVTWYKDGKLSKSSSKVHMEASGYTRRLVVBQQAQA	DAGEYSCGAGQRLSFLHVAELEPQISERPCCRREPLVVKHEHDIILTATLATPSAAT	VTLKDGVEIRRSKRHETASQGDTHLTIVHGAOVLDSAIYSCRVGAEQDFFVQVEEV	AAKFCRLLEPVCGLGTGTVTLACELSPACNAEVVMRCGNTQLRVGKRPQMVASGPPVRS	TVLGLRAEDAGEYVCRRDHDTHSAQLTVSVPRVKFMGSLTVVAEEGEATPQCVCVS	PSDVAVVFRDGLLQPSKFAISQSGASHSLTISDLVLEDAGQITVEAGASSAAEL	RVRAPVLFKKLEPQTVERRSVTLEVELTRAWPELRWTRNATALAPGKNVEIHAEAG	ARHLVLHNVGFADRGFFGCGCTPDDTKTQAKLTVE"	<2201..2468	/gene="OBSCN"	/note="C"	2469..3518	/gene="OBSCN"	/number=17	3519..3794	/gene="OBSCN"	/number=18	3795..4016	/gene="OBSCN"	/number=18	4017..4295	/gene="OBSCN"	/number=19	4296..4424	/gene="OBSCN"	/number=19	4425..4691	/gene="OBSCN"	/number=20	4692..4818	/gene="OBSCN"	/number=20	4819..5088	/gene="OBSCN"	/number=21	5089..5997	/gene="OBSCN"	/number=21	5998..>6263	/gene="OBSCN"	/number=22	ORIGIN	
Query Match		22.0%; Score 220; DB 9; Length 6400;		Best Local Similarity		85.9%; Pred. No. 1.9e-27;		Matches 244; Conservative		0; Mismatches 40; Indels 0; Gaps 0;		Qy 647		CGTCTCTGAGCCCAAGGTGGTGTGTTGCTTAAGGAGCAGCTGGCACCGAGAGCTCAGGC	706	Db 3511		CATCCAGAGCCCAAGGTGGTGTGTTGCCAAGGAGCAGCGCGCATGTCAGGAGGTGCGGC	3570	Qy 707		AGAGCGAGGAGCAGGTGCCACACTGAGCTCGAGGTGGCCAGGCCACAGCGAGGTGAC	766	Db 3571		GGAGCGGGGGCCAGTGGCACCGCTGAGCTGTGAGGTGGGCCAGGCCAGATGGAGGTGAC	3630	Qy 767		GTGGTACAAGATGGGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAGG	826	Db 3631		ATGGTACAAGACCGGAAGAAGCTGAGCTCCAGCTCAAAAGTGCACTGGAGCCGAGCGG	3690																					

gene	2201..6263				
mRNA	/gene="OBSCN"				
	join(<2201..2468,3519..3794,4017..4295,4425..4691,4819..5088,5998..>6263)				
	/gene="OBSCN"				
	join(<2201..2468,3519..3794,4017..4295,4425..4691,4819..5088,5998..>6263)				
CDS	/gene="OBSCN"				
	/codon_start=1				
	/product="obscurin"				
	/protein_id="CAC85750.1"				
	/db_xref="GI:21104334"				
	/db_xref="UniProt/TREMBL:O6NH5"				
	/translations="VVFADQPVHREVQAEAGTSTMLSCEVAQAOTEVVMYKDGKLS				
	SSSKMRVEAVGCTTRLVVQEQADAGEYSCAAGQRLSFHLHVAEPKVVFAKEQPAC				
	REVQAEAGASATLSCVQAQOMEVTVYKDGKLSKSSKVHMEASGVTTRLLVVQQAQA				
	DAGEYCEAGGQRLSPHLHVAELEPQISERPCREPLVVKHEHDIILTATLTPSAAT				
	VTWLKDGVEIRRSKRHETASQDTHLTIVHGAQVLDSAIYSCRVAEGQDPFVQVEEV				
	AAKFCRLLEPVCVELGCTVTLACELSPACAEVVMRCGNTOLRVKRPQMVASGPVRS				
	TVLGLRAEDAGEYVCESRDDHTSAQLTVSVPRVVKFMSGLSTVVAEAGEATPQCIVS				
	PSDVAVVFRDQALLQPSSEKFAISQSGASHSLITIDLVEDAGQITVEAGASSAAL				
	RVREAPVLFKKLEPQTVERRSSVTLEVLTRAMPPELWTRNATALAPGNKVEIHAEAG				
	ARHRLVILHNVPFADRGFFGCTETPDDKTQAKLTVE"				
exon	<2201..2468				
	/gene="OBSCN"				
	/note="C"				
intron	2469..3518				
	/gene="OBSCN"				
	/number=17				
exon	3519..3794				
	/gene="OBSCN"				
	/number=18				
intron	3795..4016				
	/gene="OBSCN"				
	/number=18				
exon	4017..4295				
	/gene="OBSCN"				
	/number=19				
intron	4296..4424				
	/gene="OBSCN"				
	/number=19				
exon	4425..4691				
	/gene="OBSCN"				
	/number=20				
intron	4692..4818				
	/gene="OBSCN"				
	/number=20				
exon	4819..5088				
	/gene="OBSCN"				
	/number=21				
intron	5089..5997				
	/gene="OBSCN"				
	/number=21				
exon	5998..>6263				
	/gene="OBSCN"				
	/number=22				
ORIGIN					
Query Match 22.0%; Score 220; DB 9; Length 6400;					
Best Local Similarity 85.9%; Pred. No. 1.9e-27;					
Matches 244; Conservative 0; Mismatches 40; Indels 0; Gaps 0;					
QY	647	CGTCTCTGAGCCCAAGTGGTGTGTTTGCCTAAGGAGCAGCTGGCACGCGAGGAGCTGCAGGC	706		
Db	3511	CATCCAGAGCCCAAGTGGTGTGTTTGCCAAAGGAGCAGCGCGCATGCGAGGAGTGCAGGC	3570		
QY	707	AGAGCAGGAGCCAGTGCACACTCAGCTGCGAGGTGCGCCAGGCCACAGCGAGGTGAC	766		
Db	3571	GGAGCGCGGGCCAGTGCACCGCTGAGCTGTGAGGTGGCCAGGGCCAGATGGAGGTGAC	3630		
QY	767	GTGGTACAGGATGGGAAGAGCTGAGCTCCAGCTCGAAGTGTGTCATGGGAGGCCACAGG	826		
Db	3631	ATGGTACAGGACGGGAAGAGCTGAGCTCCAGCTCAAAATGCACTGGAGGCCAGCGG	3690		

Qy	827	CTGCACCGCAGGCTGGTTGTGCACAGCAGCGCAGCGATCCGGGAGATAGCTG	886
Db	3691	CTACACGGAGGCTGTGTAGTGCACAGCGGGCGCGGATGCTCGGAGTACAGCTG	3750
Qy	887	CGAGGCTGGGGGCGCAGCGGCTCTCCTTCCATCTGGATGTCAAAG	930
Db	3751	TGAGGCGGGGGCGCACGGGTGTCTTTCGCGCTCAGCTGGCAG	3794

Search completed: March 21, 2005, 05:34:58
Job time : 4518.62 secs

**THI-SIGHT
EAGLE BLANK (USPTO)**


```
* 1225 3032: contig of 1808 bp in length
* 3033 3132: gap of unknown length
* 3133 4493: contig of 1361 bp in length
* 4494 4593: gap of unknown length
* 4594 5860: contig of 1267 bp in length
* 5861 5960: gap of unknown length
* 5961 7670: contig of 1710 bp in length
* 7671 7770: gap of unknown length
* 7771 9669: contig of 1899 bp in length
* 9670 9769: gap of unknown length
* 9770 11715: contig of 1946 bp in length
* 11716 11815: gap of unknown length
* 11816 14243: contig of 2428 bp in length
* 14244 14343: gap of unknown length
* 14344 16887: contig of 2544 bp in length
* 16888 16987: gap of unknown length
* 16988 19247: contig of 2260 bp in length
* 19248 19347: gap of unknown length
* 19348 21375: contig of 2028 bp in length
* 21376 21475: gap of unknown length
* 21476 25025: contig of 3550 bp in length
* 25026 25125: gap of unknown length
* 25126 28051: contig of 2926 bp in length
* 28052 32054: contig of 3903 bp in length
* 32055 32154: gap of unknown length
* 32155 35716: contig of 3562 bp in length
* 35717 35816: gap of unknown length
* 35817 40891: contig of 5075 bp in length
* 40892 40991: gap of unknown length
* 40992 44027: contig of 3036 bp in length
* 44028 44127: gap of unknown length
* 44128 48999: contig of 4772 bp in length
* 48900 48999: gap of unknown length
* 49000 54255: contig of 5256 bp in length
* 54256 54355: gap of unknown length
* 54356 58994: contig of 4639 bp in length
* 58995 59094: gap of unknown length
* 59095 64390: contig of 5296 bp in length
* 64391 64490: gap of unknown length
* 64491 70865: contig of 6375 bp in length
* 70866 70965: gap of unknown length
* 70966 78667: contig of 7702 bp in length
* 78668 78767: gap of unknown length
* 78768 87207: contig of 8440 bp in length
* 87208 87307: gap of unknown length
* 87308 97858: contig of 10551 bp in length
* 97859 97958: gap of unknown length
* 97959 109520: contig of 11562 bp in length
* 109521 109621: gap of unknown length
* 109622 120908: contig of 11288 bp in length
* 120909 121008: gap of unknown length
* 121009 141477: contig of 20469 bp in length
* 141478 141577: gap of unknown length
* 141578 162672: contig of 21095 bp in length
* 162673 162773: gap of unknown length
* 162773 163974: contig of 1202 bp in length
* 163975 164075: gap of unknown length
* 164075 164766: contig of 692 bp in length.
```

FEATURES

Source

```
1..164766
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-245P10"
1..1124
/note="assembly_name:Contig14"
1225..3032
/note="assembly_name:Contig21"
3133..4493
/note="assembly_name:Contig23"
4594..5860
misc_feature
misc_feature
misc_feature
misc_feature
```

```
misc_feature
/note="assembly_name:Contig24"
5961..7670
/note="assembly_name:Contig25"
7771..9669
/note="assembly_name:Contig26"
9770..11715
/note="assembly_name:Contig27"
11816..14243
/note="assembly_name:Contig28"
14344..16887
/note="assembly_name:Contig29"
16988..19247
/note="assembly_name:Contig30"
19348..21375
/note="assembly_name:Contig31"
21476..25025
/note="assembly_name:Contig32"
25126..28051
/note="assembly_name:Contig33"
28152..32054
/note="assembly_name:Contig34"
32155..35716
/note="assembly_name:Contig35"
35817..40891
/note="assembly_name:Contig36"
40992..44027
/note="assembly_name:Contig37"
44128..48999
/note="assembly_name:Contig38"
49000..54255
/note="assembly_name:Contig39"
54356..58994
/note="assembly_name:Contig40"
59095..64390
/note="assembly_name:Contig41"
64491..70865
/note="assembly_name:Contig42"
70966..78667
/note="assembly_name:Contig43"
78768..87207
/note="assembly_name:Contig44"
87308..97858
/note="assembly_name:Contig45"
97959..109520
/note="assembly_name:Contig46"
109621..120908
/note="assembly_name:Contig47"
121009..141477
/note="assembly_name:Contig48"
141578..162672
/note="assembly_name:Contig49"
162773..163974
/note="assembly_name:Contig9"
164075..164766
/note="assembly_name:Contig7"
```

ORIGIN

```
Query Match      59.8%; Score 239.8; DB 2; Length 164766;
Best Local Similarity 99.2%; Pred. No.1e-38;
Matches 241; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GAAGACCAAGTGGTGGCGCCAGGGAGGACGTGGAGCTGCCTGTGAGCTGTACCGGCG 60
Db 60103 GAAGACCAAGTGGTGGCGCCAGGGAGGACGTGGAGCTGTGAGCTGTACCGGCG 60162

Oy 61 GGAAGCCCGTGCACCTGGCTGAAGACAGGAGGACATCCGCAAGAGCCAGAAGTATGAT 120
Db 60163 GGAAGCCCGTGCACCTGGCTGAAGACAGGAGGACATCCGCAAGAGCCAGAAGTATGAT 60222

Oy 121 GTGGTCTCGAGGGCAGCATGGCCATGCTGTGCTATCCCGGGGGCTCGCTCAAGACGCG 180
Db 60223 GTGGTCTCGAGGGCAGCATGGCCATGCTGTGCTATCCCGGGGGCTCGCTCAAGACGCG 60282
```

QY 181 GGCAGTACAGTGTGAGTGGAGGCTTCCAGAGACACAGCCAGCTCCATGTGGAGAA 240
Db 60383 GGCAGTACAGTGTGAGTGGAGGCTTCCAGAGACACAGCCAGCTCCATGTGGAGGT 60342
QY 241 AAA 243
Db 60343 AAA 60345

RESULT 11
AC023889
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
ACCESSION AC023889
VERSION AC023889.3 GI:8969253
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE Waterston,R.H.
AUTHORS The sequence of Homo sapiens clone
TITLE Unpublished
REFERENCE 2 (bases 1 to 174612)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 7, 2000 this sequence version replaced gi:8748947.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----
Center project name: H_NH0861B12
----- Summary Statistics -----

Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2478: contig of 2478 bp in length
* 2479 2578: gap of unknown length
* 2579 6948: contig of 4370 bp in length
* 6949 7048: gap of unknown length
* 7049 9373: contig of 2325 bp in length
* 9374 9474: gap of unknown length
* 9474 13303: contig of 3830 bp in length
* 13304 13403: gap of unknown length
* 13404 18270: contig of 4867 bp in length
* 18271 25589: gap of unknown length
* 25590 25689: gap of unknown length

* 25690 34815: contig of 9126 bp in length
* 34816 34915: gap of unknown length
* 34916 45742: contig of 10827 bp in length
* 45743 45842: gap of unknown length
* 45843 62987: contig of 17145 bp in length
* 62988 63087: gap of unknown length
* 63088 87584: contig of 24497 bp in length
* 87585 87684: gap of unknown length
* 87685 114765: contig of 27081 bp in length
* 114766 114865: gap of unknown length
* 114866 142334: contig of 27469 bp in length
* 142335 142435: gap of unknown length
* 142435 174612: contig of 32178 bp in length.

FEATURES
Location/Qualifiers
1..174612
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-661B12"

misc_feature 1..2478
/note="assembly_name:Contig4"
misc_feature 2579..6948
/note="assembly_name:Contig5"
misc_feature 7049..9373
/note="assembly_name:Contig6"
misc_feature 9474..13303
/note="assembly_name:Contig7"
misc_feature 13404..18270
/note="assembly_name:Contig8"
misc_feature 18371..25589
/note="assembly_name:Contig9"
misc_feature 25690..34815
/note="assembly_name:Contig10"
misc_feature 34916..45742
/note="assembly_name:Contig11"
misc_feature 45843..62987
/note="assembly_name:Contig12"
misc_feature clone end:SP6
vector side:right
misc_feature 63088..87584
/note="assembly_name:Contig13"
misc_feature 87685..114765
/note="assembly_name:Contig14"
misc_feature 114866..142334
/note="assembly_name:Contig15"
misc_feature 142435..174612
/note="assembly_name:Contig16"

ORIGIN

Query Match 59.8%; Score 239.8; DB 2; Length 174612;
Best Local Similarity 99.2%; Pred. No. 1e-38;
Matches 241; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GAAGACCACTGGTGGCGCCAGGAGGACGTGGAGCTGCTGTGAGCTGTACCGGCG 60
Db 110510 GAAGACCACTGGTGGCGCCAGGAGGACGTGGAGCTGCTGTGAGCTGTACCGGCG 110569
QY 61 GGAAGCCCGTGCACCTGGCTGAAGGACAGGAGGCCATCCGCAAGACGCAAGTATGAT 120
Db 110570 GGAAGCCCGTGCACCTGGCTGAAGGACAGGAGGCCATCCGCAAGACGCAAGTATGAT 110629
QY 121 GTGGTCTCGAGGGCACCATGGCCATGCTGGTTCATCCGCGGGGCCCTCGCTCAAGGACGCG 180
Db 110630 GTGGTCTCGAGGGCACCATGGCCATGCTGGTTCATCCGCGGGGCCCTCGCTCAAGGACGCG 110689
QY 181 GCGAGTACAGTGTGAGTGGAGGCTTCCAGAGACAGCCAGCCTCCATGTGGAAGAA 240
Db 110690 GCGAGTACAGTGTGAGTGGAGGCTTCCAGAGACAGCCAGCCTCCATGTGGAAGGT 110749
QY 241 AAA 243
Db 110750 AAA 110752

RESULT 12
AL645854/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
COMMENT

AL645854
Mouse DNA sequence from clone RP23-192P17 on chromosome 11,
complete sequence.
AL645854
AL645854.10 GI:20068636
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Smith, M.
Direct Submission
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:19699553.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: ENBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-192P17 is
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6.
Location/Qualifiers
1. 211829
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-192P17"
/clone_lib="RPCI-23"

ORIGIN
Query Match 35.6%; Score 142.8; DB 10; Length 211829;
Best Local Similarity 74.4%; Pred. No. 3.2e-19;
Matches 180; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 GAAGACCACTGGTGGCGCCAGGAGGACGTGGAGCTGCGCTGTGAGCTGTCCAGCGCG 60
DB 57090 GAGGACCACTGGGACGACGCTTGGGAGAGATGTAAGTCTGAGCTGTCCAGGCA 57031
QY 61 GGAACCGCGCTGCTGCTGAAGGACAGGAGGCGCATCCGCAAGGCGCAGAGTATGAT 120
DB 57030 GGCACCTCTGTGCGCTGGCTGAAGACGGAAGGCGCATCAGGAGAGCAGAGTATGAC 56971
QY 121 GTGCTTCGAGGCGCAGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
DB 56970 CTGCTCAGTGAAGGCGCGGCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 56911
QY 181 GCGGATACCGTGTGAGGTGGAGGCTTCCAGAGCAGCAGCGCTCCATGTGGAAGAA 240
DB 56910 GGGGATACCTTGGAGACAGAGCTTCCAGAGCAGTGCACAGCTCTGTGTGAGGT 56851

QY 241 AA 242
DB 56850 AA 56849

RESULT 13
AC099089/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
COMMENT

AC099089
Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.
AC099089
AC099089.5 GI:30522701
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 260998)
Murny D.Marie, Metzker M.Lee, Abramson S., Adams C., Alder J.,
Allen C., Allen H., Albrooks S., Amin A., Anguiano D.,
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed P.,
Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Cleaveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
Delgado M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
Delgado O., Denison S., Detamo C., Ding Y., Dinh H., Divya K.,
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
Georgiev G.E., Geer K., Gill R., Grady M., Guerra W., Guevara W.,
Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K.,
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
Hernandez R., Hines S., Hladun S.L., Hodgson A., Hognes M.,
Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Karpachy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
Kovis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
Lorenshewa L., Louised H., Lozano R.J., Lu X., Ma J.,
Maheshwari M., Mahindratne M., Mahmoud M., Malloy K., Mangum A.,
Mangum B., Mapua P., Martin K., Martin R., Martin E.,
Mawhinney S., McLeod M.P., McNeill T.Z., Meenen E.,
Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,
Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,
Nankervis C., Neal D., Newton N., Nguyen N., Norris S., Parks K.,
Nwackemehe O., Okwuonu G., Olarnpunsagoon A., Pal S., Parks K.,
Plopper F., Poindexter A., Popovic D., Primus E., Pfannkuch C.,
Puzo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
Rives C., Rokey T., Rojas A., Rose M., Rose R., Ruiz S., Shen H.,
Shetty J., Shvartsbeyn A., Sison I., Sitter C.D., Smaja D.,
Snead J., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,
Steinle M., Strong R., Sutton A., Svatek A., Tabor P., Taylor C.,
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Umant K.,
Valas R., Vera V., Villasana D., Waldron L., Walker B., Wang J.,
Wang Q., Wang S., Warren J., Warren R., Wei X., White F.,
Williams G., Willson R., Wleczyk R., Wooden H., Worley K.,
Wright D., Wright J., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
Yu F., Zhang J., Zhou X., Zhou S., Zhao S., Dunn D., von
Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,
Weinstock G. and Gibbs R.A.
Unpublished
2 (bases 1 to 260998)
Worley K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department


```

Assembly program: Phrap; version 0.990319
Consensus quality: 97841 bases at least Q40
Consensus quality: 98004 bases at least Q30
Consensus quality: 98090 bases at least Q20
Insert size: 85000; agarose-fp
Insert size: 98113; sum-of-contigs
Quality coverage: 11.51x in Q20 bases; agarose-fp
Quality coverage: 9.97x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 28467: contig of 28467 bp in length
28468 28567: gap of unknown length
28568 31135: contig of 2568 bp in length
31136 31235: gap of unknown length
31236 41487: contig of 10252 bp in length
41488 41587: gap of unknown length
41588 74365: contig of 32778 bp in length
74366 74466: gap of unknown length
74466 96541: contig of 22076 bp in length
96542 96641: gap of unknown length
96642 98613: contig of 1972 bp in length.
-----
Location/Qualifiers
1. 98613
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="221D8"
/clone_lib="Incyte Genomics"
misc_feature
1. 28467
/note="assembly_fragment
clone_end:T7
vector_side:left"
28568..31135
/note="assembly_fragment"
31236..41487
/note="assembly_fragment"
41588..74365
/note="assembly_fragment"
74466..96541
/note="assembly_fragment"
96642..98613
/note="assembly_fragment
clone_end:SP6
vector_side:right"
ORIGIN
Query Match 17.3%; Score 69.2; DB 2; Length 98613;
Best Local Similarity 59.8%; Pred. No. 0.00024;
Matches 116; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
Qy 12 GGTGGCGGAGGAGGACGTGGAGCTGGCTGTGAGCTGTACCGGGCGGGAACGCCCGT 71
Db 57893 GGCAGTCCGGCGGAGGACGTGGAGCTGTGAGATCACCAGACGGAGTCACCAT 57952
Qy 72 GCATGGCTGAAGGACAGGAAGGCCATCCGCAAGAGCCAGAGTATGATGTGCTGCGA 131
Db 57953 TCGGTGGCTGAAATATGCCATTTGATCCGGCAGAGCCCAAAGTATGATGATGTTC 58012
Qy 132 GGGCAGATGGCCATGCTGGTCATCCGGGGGCTCGCTCAAGGCGGGCGGAGTACAC 191
Db 58013 GAAGAACAGGCGGAGCTGGTGTATCAGGAACGATCAGGACTCTGGAGATTACTG 58072
Qy 192 GTGTGAGGTGGAGG 205
Db 58073 CTGGAGGACAGATG 58086

```

```

RESULT 15
AC023889/c
LOCUS
DEFINITION
AC023889 Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION
AC023889.3 GI:8969253
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 174612)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 174612)
Waterston,R.H.
Direct Submission
Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jul 7, 2000 this sequence version replaced gi:8748947.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
-----
Center project name: H_NH0661B12
-----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 2478: contig of 2478 bp in length
2479 2578: gap of unknown length
2579 6948: contig of 4370 bp in length
6949 7049: gap of unknown length
7049 9373: contig of 2325 bp in length
9374 9474: gap of unknown length
9474 13303: contig of 3830 bp in length
13304 13403: gap of unknown length
13404 18270: contig of 4867 bp in length
18271 18371: gap of unknown length
18371 25590: contig of 7219 bp in length
25590 25690: gap of unknown length
25690 34816: contig of 9126 bp in length
34816 34916: gap of unknown length
34916 45743: contig of 10827 bp in length
45743 45842: gap of unknown length
45843 62987: contig of 17145 bp in length
62988 63087: gap of unknown length
63088 87584: contig of 24497 bp in length

```

```

* 87585 87684: gap of unknown length
* 87685 114765: contig of 27081 bp in length
* 114766 114865: gap of unknown length
* 114866 142334: contig of 27469 bp in length
* 142335 142434: gap of unknown length
* 142435 174612: contig of 32178 bp in length.

```

FEATURES

source

```

1. .174612
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="1"
/clone="RP11-661B12"
1. .2478
/note="assembly_name:Contig4"
misc_feature
2579. .6948
/note="assembly_name:Contig5"
misc_feature
7049. .9373
/note="assembly_name:Contig6"
misc_feature
9474. .13303
/note="assembly_name:Contig7"
misc_feature
13404. .18270
/note="assembly_name:Contig8"
misc_feature
18371. .25589
/note="assembly_name:Contig9"
misc_feature
25690. .34815
/note="assembly_name:Contig10"
misc_feature
34916. .45742
/note="assembly_name:Contig11"
misc_feature
45843. .62987
/note="assembly_name:Contig12
clone_end:SP6
vector_side:right"
63088. .87584
/note="assembly_name:Contig13"
misc_feature
87685. .114765
/note="assembly_name:Contig14"
misc_feature
114866. .142334
/note="assembly_name:Contig15"
142435. .174612
/note="assembly_name:Contig16"

```

ORIGIN

```

Query Match          17.2%; Score 69; DB 2; Length 174612;
Best Local Similarity 56.3%; Fred. NO. 0.00024;
Matches 129; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 5 ACCAGTGGGTGGCGCCAGGGGAGGACGTGGAGCTGCGCTGTGAGCTGTCAACGGCGGGAA 64
Db 155238 ACCAGGAGGTGGATGAGGGAGGCCGACACTTATGCTGCGAGCTGAGCCGGCGGGTG 155179

QY 65 CGCCCGTCACCTGGCTGAAGACAGGAGGCCATCCCGCAAGAGCCAGAGTATGATGTGG 124
Db 155178 CGAGCGTGGAGTGGCGCAAGGGCTCCCTACAGCTCTTCCCTTGTGCCAAGTACCAGATGG 155119

QY 125 TCTCGAGGGCACCAGATGGCCATGCTGTGTCATCCGGGGGCTCGCTCAAGACGCGGGCG 184
Db 155118 TGCAGGATGGTGCAGCTGCAGAGTGTGTTGACCGGAGTGGAGCAGGAGATGCGGGTG 155059

QY 185 AGTACACGTGTGAGGTGAGGGCTTCCAAGACACAGCCAGCCTCCATGT 233
Db 155058 ACTACACGTGTGACACGGGCCACACGACGAGCATGGCCAGCCTCTCTGT 155010

```

Search completed: March 21, 2005, 05:35:02
 Job time : 1812.95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:24:10 ; Search time 4966.73 Seconds
(without alignments)
10741.310 Million cell updates/sec

Title: US-10-077-130-6_COPY_13600_14700

Perfect score: 1101
Sequence: 1 cacagcagcacactgtgac.....agcagggaaggcccatgttc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_ats.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1101	100.0	20435	9 HSAJ2535	AJ002535 Homo sapi
2	384	34.9	18760	9 HSAJ4906	AJ314906 Homo sapi
3	384	34.9	135964	9 AL353593	AL353593 Human DNA
C 4	380.8	34.6	174612	2 AC023889	AC023889 Homo sapi
5	350.8	31.9	164766	2 AC026657	AC026657 Homo sapi
C 6	277.8	25.2	260998	2 AC099089	AC099089 Rattus no
C 7	267.6	24.3	211829	10 AL645854	AL645854 Mouse DNA
8	248.4	22.6	18524	6 CQ730656	CQ730656 Sequence
9	200.8	18.2	174612	2 AC023889	AC023889 Homo sapi
C 10	75.4	6.8	168144	2 CR792456	CR792456 Danio rer
C 11	74.8	6.8	125020	2 AF429315	AF429315 Homo sapi
12	71.2	6.5	2000	6 AX655393	AX655393 Sequence
13	58.4	5.3	125020	9 AF429315	AF429315 Homo sapi
14	53.4	4.9	2534	6 AX834822	AX834822 Sequence
15	53.4	4.9	2534	9 AK097489	AK097489 Homo sapi
16	53.4	4.9	2761	9 BC043916	BC043916 Homo sapi
17	53.4	4.9	4789	9 AB046776	AB046776 Homo sapi
18	53.4	4.9	34667	9 HSAJ14905	AJ314905 Homo sapi
19	50.4	4.6	80182	5 AL772356	AL772356 Zebrafish

C 20	50.4	4.6	194266	2 BX571737	BX571737 Danio rer
C 21	50.2	4.6	1005	6 CQ445545	CQ445545 Sequence
C 22	49.4	4.5	2000	6 AX655393	AX655393 Sequence
C 23	49.4	4.5	31828	2 AC151613	AC151613 Emiliania
C 24	49.4	4.5	157754	10 AC132313	AC132313 Mus muscu
25	48.6	4.4	1772	9 HUMVBPBA	L05606 Human myosi
26	48.6	4.4	1775	6 CQ721696	CQ721696 Sequence
27	48.6	4.4	1821	9 BC044226	BC044226 Homo sapi
28	47.6	4.3	5390	9 HSAJ14904	AJ314904 Homo sapi
29	47.4	4.3	5992	10 BC058610	BC058610 Mus muscu
30	47.4	4.3	6021	10 AY237726	AY237726 Mus muscu
31	47.2	4.3	197411	10 AC135961	AC135961 Mus muscu
32	47	4.3	152618	10 AC127583	AC127583 Mus muscu
33	46.6	4.2	3613	9 HSM802220	AL137493 Homo sapi
34	46.6	4.2	3951	9 AY245430	AY245430 Homo sapi
35	46	4.2	27060	3 AB055927	AB055927 Procamb
36	46	4.2	91501	2 AC024344	AC024344 Homo sapi
C 37	46	4.2	147499	2 AY178788	AY178788 Canle fam
C 38	46	4.2	277000	1 SCO93109	AL939109 Streptomy
C 39	45.6	4.1	648	12 AY657521	AY657521 Synthetic
40	45.6	4.1	1688	8 AK071050	AK071050 Oryza sat
41	45.6	4.1	1791	8 AK105948	AK105948 Oryza sat
C 42	45.6	4.1	2010	6 E01992	E01992 DNA sequenc
43	45.6	4.1	12421	1 AE004881	AE004881 Pseudomon
C 44	45.6	4.1	110000	1 AP006618_33	Continuation (34 o
C 45	45.6	4.1	141079	8 AP002868	AP002868 Oryza sat

ALIGNMENTS

RESULT 1	HSAJ2535	Homo sapiens mRNA for obscurin (OBSCN gene).	20435 bp	linear	PRI 14-SEP-2001
LOCUS	AJ002535	OBSCN gene; obscurin.			
ACCESSION	AJ002535.1	GI:15026973			
VERSION	OBSCN gene; obscurin.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	Young, P., Ehler, E. and Gautel, M.				
AUTHORS	Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly				
TITLE	J. Cell Biol. 154 (1), 123-136 (2001)				
JOURNAL	21342081				
MEDLINE	11448995				
PUBMED	2 (bases 1 to 20435)				
REFERENCE	Gautel, M.S.				
AUTHORS	Direct Submission				
TITLE	Submitted (31-OCT-1997) Gautel M.S., Structural Biology Division, European Molecular Biology Laboratory, Meierhofstr. 1, Heidelberg, 69117, GERMANY				
JOURNAL	Revised by author 20-JUL-2001				
REMARK	Location/Qualifiers				
FEATURES	1..20435				
source	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/map="1q42.1"				
	/cell_type="cardiomyocyte"				
	/tissue_type="cardiac"				
	/dev_stage="adult"				
gene	1..20435				
	/gene="OBSCN"				
CDS	72..19934				
	/gene="OBSCN"				
	/function="titin binding"				
	/codon_start=1				
	/evidence=experimental				

Qy	301	CAGGTGGCAGCTGGTGAAGATGCTCTCTCTGGAGCTTGAGTGGTGGCTGAGGCTGGTGAAG	360
Db	13971		
Qy	361	GTCTATCTGGCACAAGGGAATGAGCGCATCAGACCCGGTGGCGGCTTCGAGTGGTCTGCC	420
Db	14031	GTCTATCTGGCACAAGGGAATGAGCGCATCAGACCCGGTGGCGGCTTCGAGTGGTCTGCC	14090
Qy	421	CAGGTGGCACAAGGGAATGAGCGCATCAGACCCGGTGGCGGCTTCGAGTGGTCTGCC	480
Db	14091		
Qy	481	CAGGTGGCACAAGGGAATGAGCGCATCAGACCCGGTGGCGGCTTCGAGTGGTCTGCC	540
Db	14151	CAGGTGGCACAAGGGAATGAGCGCATCAGACCCGGTGGCGGCTTCGAGTGGTCTGCC	14210
Qy	541	AGCCGAGGCTCTGTGGATGAGCGCCCTCAGCGCCAGCTTTGCCCGCCGAGCGCAGCCAGGAG	600
Db	14211	AGCCGAGGCTCTGTGGATGAGCGCCCTCAGCGCCAGCTTTGCCCGCCGAGCGCAGCCAGGAG	14270
Qy	601	GGTGACCTGCACCTACTGTGGAGGCGCTGGCTCGGAAACGTGCGATGAGCGGTGAGGCC	660
Db	14271	GGTGACCTGCACCTACTGTGGAGGCGCTGGCTCGGAAACGTGCGATGAGCGGTGAGGCC	14330
Qy	661	ACGCTGGACTCCATTAGCGAGCTGCAGAGGAGCGCGCTCGCAGCGGCTGCCACAG	720
Db	14331	ACGCTGGACTCCATTAGCGAGCTGCAGAGGAGCGCGCTCGCAGCGGCTGCCACAG	14390
Qy	721	GAGCGAGGAGGCTGGACCTGATCTCTGAAGGCTACTCCAGCGCGATGAGCTGGCC	780
Db	14391	GAGCGAGGAGGCTGGACCTGATCTCTGAAGGCTACTCCAGCGCGATGAGCTGGCC	14450
Qy	781	CGCACTGAGATGCTGACCTCTCACACACAGCTCTGATGATGAGTCCCGGCGAGGACCC	840
Db	14451	CGCACTGAGATGCTGACCTCTCACACACAGCTCTGATGATGAGTCCCGGCGAGGACCC	14510
Qy	841	CCTTCCCTGGTCACTACTCAAGAGGCTGGAGCGCAGGACCTCACCCTGCGCAGC	900
Db	14511	CCTTCCCTGGTCACTACTCAAGAGGCTGGAGCGCAGGACCTCACCCTGCGCAGC	14570
Qy	901	AAGGTGGGGGCCCCAGCAGCCCTCTCTGAAGCCACAGCAGCAGCAGGACCACTGGCT	960
Db	14571	AAGGTGGGGGCCCCAGCAGCCCTCTCTGAAGCCACAGCAGCAGCAGGACCACTGGCT	14630
Qy	961	GCTGTGCCCCACCACTGGAGACTGAGCAGCACCAGACCTGGGTGATCCCTCAATGAC	1020
Db	14631	GCTGTGCCCCACCACTGGAGACTGAGCAGCACCAGACCTGGGTGATCCCTCAATGAC	14690
Qy	1021	AAGCAGCTGTGAGATCCAGCTGCCCTTAAAGGCTACAGGTCGGAAGGATGAAG	1080
Db	14691	AAGCAGCTGTGAGATCCAGCTGCCCTTAAAGGCTACAGGTCGGAAGGATGAAG	14750
Qy	1081	CAGCAGGAAGGGCCCATGTTTC	1101
Db	14751		

RESULT 2	HS314906	18760 bp	DNA	linear	PRI 21-MAY-2002
LOCUS	HS314906				
DEFINITION	Homo sapiens partial OBSCN gene for obscurin, exons 57-68.				
ACCESSION	AJ314906				
VERSION	AJ314906.1	GI:21104339			
KEYWORDS	OBSCN gene; obscurin.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Young, P., Ehler, E. and Gautel, M.				
TITLE	Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly				
JOURNAL	Unpublished				

REFERENCE	2 (bases 1 to 18760)				
AUTHORS	Gautel, M.S.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie, Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse 11, Dortmund, 44227, GERMANY				
FEATURES	Location/Qualifiers				
source	1..18760				
gene	/organism="Homo sapiens"				
mRNA	/mol_type="genomic DNA"				
CDS	/db_xref="taxon:9606"				
exon	149..18738				
intron	/gene="OBSCN"				
exon	join<149..424,904..1191,1332..1613,2324..2694,4784..5654,7668..7046,16300..16377,16635..16763,17023..17220,17952..17984,18255..18353,18520..>18738)				
intron	/gene="OBSCN"				
exon	join<149..424,904..1191,1332..1613,2324..2694,4784..5654,7668..7046,16300..16377,16635..16763,17023..17220,17952..17984,18255..18353,18520..>18738)				
intron	/gene="OBSCN"				
exon	/codon_start=3				
intron	/product="obscurin"				
exon	/protein_id="CAC85753.1"				
intron	/db_xref="GI:21104340"				
exon	/db_xref="UniProt/TREMBL:Q8NH2"				
intron	/translation="WRLEILPLKNAAVRAGAQAQCTCTLSEA VVPGASWINGAAV QPDDSDWTADGSHHALLRLSRQPHHAGEVTPACRDVAVASARLTVLGLPPDEDAEV VARSHVTVTLSWAAPMSDGGGCGCYVEVKEGATGWRLCHELVPGEVVDGLAPG ETVFRVAAGVPGVAGRPVHLPTQVRLAEPKPPVPOPSAPESQVAAEGBVSLLELV VAEGEVIWHKGMERI QPGCFEFVVSQGRQOMLVIKGTAEADQGEYHCGLAQGSICPA AATQVALSPASVDEAFQPSLPPAAEGDHLLEALARKRMSREPTLDSISLPE EDGRSRLPQAEAEVAPDLSEGYSTADELARTADLSTSSDDESRAGTPSLVTLK KAGPFGTSPASKVGAAPAAKVPQQQQELAAVRLPLGDLSTKDLGDFSMDKAAVKI QAAFGYKVRKEMKQEQPMPSHTFGDTEAOVGDALRLCEVVASKADVRARWLKDGVE LTRGRHHIDOLGDTCSLLITGLDRADAGCTCOVSNKFGVTHSACVVVSGSSEA ESSGGELDDAFRAARLRHLRFLTKSPAESVDELEFLSADGPAEPREPADMQTYRE DEHFCIRFEALTEARQAVTRFQMFATLGVSEIKLVEQGPFRVEMCISKETPAPVV PPEPILLSLTDAAVFLTELQNDQVQDGPVDFVVTGQPMPSVRFMDKGLLEED DHYINEDQGGHQI1ITAVVADMGVYRCIAENSMGVSSTKABLRVLDTLTDYDQAA DYNSSGVFAQGYLSTSRQEGESTTDEQQLPVVBEELDLQVAPQATRIAKFLKV KYGPAPRLYKFDQGPILTASAHIRMTDKKILHTLEIISVTRKEDSGQAYIISNMGAA KYSARLLVRGDEPEEPKPSADVHEQLVPPMLERFTPKKVKKGSITFSVKEGRVVP TVHWRBEARGVLWIGDTPGTYTVASSAQOQSHLVLLDVGQRHQSTYTCIASNAAGQA LCSASLHVSG"				
exon	<149..424				
intron	/gene="OBSCN"				
exon	/number=57				
intron	425..903				
exon	/gene="OBSCN"				
intron	/number=57				
exon	904..1191				
intron	/gene="OBSCN"				
exon	/number=58				
intron	1192..1331				
exon	/gene="OBSCN"				
intron	/number=58				
exon	1332..1613				
intron	/gene="OBSCN"				
exon	/number=59				
intron	1614..2323				
exon	/gene="OBSCN"				
intron	/number=59				
exon	2324..2694				
intron	/gene="OBSCN"				
exon	/number=60				
intron	2695..4783				
exon	/gene="OBSCN"				
intron	/number=60				
exon	4784..5654				
intron	/gene="OBSCN"				
exon	/number=61				
intron	5655..6767				

		/gene="OBSCN" /number=61 6768..7046 /gene="OBSCN" /number=62 7047..16299 /gene="OBSCN" /number=62 16300..16377 /gene="OBSCN" /number=63 16378..16634 /gene="OBSCN" /number=63 16635..16763 /gene="OBSCN" /number=64 16764..17022 /gene="OBSCN" /number=64 17023..17220 /gene="OBSCN" /number=65 17221..17951 /gene="OBSCN" /number=65 17952..17984 /gene="OBSCN" /number=66 17985..18254 /gene="OBSCN" /number=66 18255..18353 /gene="OBSCN" /number=67 18354..18519 /gene="OBSCN" /number=67 18520..>18738 /gene="OBSCN" /number=68	ORIGIN	
Query Match 34.9%; Score 384; DB 9; Length 18760; Best Local Similarity 78.3%; Pred. No. 3.1e-61; Matches 540; Conservative 0; Mismatches 10; Indels 140; Gaps 1;				
QY	1	CACAGCAGCCACTGTGCACACTGTCTTGGGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGCTGTGCCAC		
Db	942	CGCAGCAGCCACTGTGCACACTGTCTTGGGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGCTGTGCCAC		
QY	61	CTCTGTGGCTACCGGCTGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGCTGTGCCAC		
Db	1002	CTCTGTGGCTACCGGCTGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGCTGTGCCAC		
QY	121	GAGCTGTGCTGACACCGAGTGTGTGGTGAATGGCTGTGGCCCGGGGAGACTTACCGC		
Db	1062	GAGCTGTGCTGACACCGAGTGTGTGGTGAATGGCTGTGGCCCGGGGAGACTTACCGC		
QY	181	TTCCGTGTGGCAGCTGTGGGCGCTGTGGGTCTGGGAAACCGGTTACCTGCCCCAGACA		
Db	1122	TTCCGTGTGGCAGCTGTGGGCGCTGTGGGTCTGGGAAACCGGTTACCTGCCCCAGACA		
QY	241	GTGGGGCTT-----249		
Db	1182	GTGGGGCTTGGTGAGTTGCTTTCATTGGGGTCTGGGGAGTTTACAGCCCCCACAAGCTGGA		
QY	250	-----249		
Db	1242	GGAGGGAGGATCCCCGGGAGTGGGGCTGCACCTGTGGGTGATCTTTGTGTATGGGGCC		
QY	250	-----GCAGAGCCACCGAAGCCTGTGCTTCCCTCCCGCAGC 		

Db	1302	TGTCTCTGTCCGTGGAGCGCTGTTCCCGAGCAGAGCCACCGAAGCCTGTGCTCCCGCAGC	1361
Qy	281	CCTCAGCCCTTGAGAGCGGCGAGGTGGCAGCTGGTGAAGATGTCTCTCTGGAGCTTGAGG	340
Db	1362	CCTCAGCCCTTGAGAGCGGCGAGGTGGCAGCTGGTGAAGATGTCTCTCTGGAGCTTGAGG	1421
Qy	341	TGCTGGCTGAGGCTGGTGAAGTCACTGGCCACAAGGGAATGGAGCGCATCAGCCCCGCTG	400
Db	1422	TGGTGGCTGAGGCTGGTGAAGTCACTGGCCACAAGGGAATGGAGCGCATCAGCCCCGCTG	1481
Qy	401	GGCGTTTCGAGGTGGTCTCCAGGGTCGGCAACAGATGCTGGTGATCAAGGGCTTCACGG	460
Db	1482	GGCGTTTCGAGGTGGTCTCCAGGGTCGGCAACAGATGCTGGTGATCAAGGGCTTCACGG	1541
Qy	461	CAGAAGACCAAGGGCGAGTCACTGTGGCTCAGGGCTCCATCTGCGCTCGGCTG	520
Db	1542	CAGAAGACCAAGGGCGAGTCACTGTGGCTCAGGGCTCCATCTGCGCTCGGCTG	1601
Qy	521	CCACCTTCCAGGTGGCACTGAGCCAGCCT	550
Db	1602	CCACCTTCCAGGTGGCGTGTGCTCCTGTCT	1631

RESULT 3
AL353593
LOCUS
DEFINITION Human DNA sequence from clone RP5-1139B12 on chromosome 1q42.1-43,
complete sequence.
ACCESSION AL353593
VERSION AL353593.33 GI:18673899
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135964)
AUTHORS Almeida, J.
TITLE Direct Submission
JOURNAL Submitted (13-PEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Feb 14, 2002 this sequence version replaced gi:17977879.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/c_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP5-1139B12 is from the library RPI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-1139B12.

FEATURES
source
1..135964
/organism="Homo sapiens"
/mol_type="genomic DNA"

```
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.1-43"
/clone="RPS-1139B12"
/clone_lib="RPCI-5"
25928..26054
/misc_feature
  /note="Sequence from overlapping clone RP11-520H14
(ALJ359510). Assembly confirmed by restriction digest."
ORIGIN
Query Match      34.9%; Score 384; DB 9; Length 135964;
Best Local Similarity 78.3%; Pred. No. 1.9e-61;
Matches 540; Conservative 0; Mismatches 10; Indels 140; Gaps 1;

QY 1 CACAGCACCACACTGTGACACTGTCTGGGAGCTCCCATGAGTATGAGGCGGTGT 60
DB 104829 CGCAGCACCACACTGTGACACTGTCTGGGAGCTCCCATGAGTATGAGGCGGTGT 104888
QY 61 CTCTGTGCTACCGGTGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGTGGCCAC 120
DB 104889 CTCTGTGCTACCGGTGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGTGGCCAC 104948
QY 121 GAGCTGTGCTGACCCGAGTGTGTGTGATGTCCTGGCCCGGGGAGACCTTACCGC 180
DB 104949 GAGCTGTGCTGACCCGAGTGTGTGTGATGTCCTGGCCCGGGGAGACCTTACCGC 105008
QY 181 TTCCGTGTGGCAGCTGTGGCCCTGTGGGTGTGGGGAAACCGGTTCACTGCCCCAGACA 240
DB 105009 TTCCGTGTGGCAGCTGTGGCCCTGTGGGTGTGGGGAAACCGGTTCACTGCCCCAGACA 105068
QY 241 GTGGGCTT----- 249
DB 105069 GTGGGCTTGTGAGTCTTCATTGGGGTCTGGGGAGTTACAGCCCCCACAGCTGGA 105128
QY 250 ----- 249
DB 105129 GGAGGGAGGATCCCGGAGTGGGGCTGCACCTGGGTGCATCTTGTGATGGGGCC 105188
QY 250 -----GCAGAGCCACCGAAGCCTGTGCTCCCGCAGC 280
DB 105189 TGTCTCTGTCCTGGAGCGCTTCTCCCGCAGAGAGCCACCGAAGCCTGTGCTCCCGCAGC 105248
QY 281 CCTAGCCCTTGAGAGCGGAGGTGGCAGCTGGTGAAGATCTCTCTGGAGCTTGAGG 340
DB 105249 CCTAGCCCTTGAGAGCGGAGGTGGCAGCTGGTGAAGATCTCTCTGGAGCTTGAGG 105308
QY 341 TGGTGGCTGAGGCTGGTGAAGTCACTTGGCCACAGGGAATGAGGCGCATCCAGCCCGGTG 400
DB 105309 TGGTGGCTGAGGCTGGTGAAGTCACTTGGCCACAGGGAATGAGGCGCATCCAGCCCGGTG 105368
QY 401 GCGGTTTCAGAGTGTCTCCAGAGGTGGGCAACAGATCTGGTATCAAGGGCTTACGG 460
DB 105369 GCGGTTTCAGAGTGTCTCCAGAGGTGGGCAACAGATCTGGTATCAAGGGCTTACGG 105428
QY 461 CAGAGACAGGGCGAGTACCACCTGTGGCTGGCTCAGGGTCCATCTGCCCTGGCGGTG 520
DB 105429 CAGAGACAGGGCGAGTACCACCTGTGGCTGGCTCAGGGTCCATCTGCCCTGGCGGTG 105488
QY 521 CCACCTTCCAGTGGCACTTGAGCCCGAGCT 550
DB 105489 CCACCTTCCAGTGGTGGTGTGTCCTGTCCT 105518

RESULT 4
AC023889/c
LOCUS AC023889 174612 bp DNA linear HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION AC023889
VERSION AC023889.3 GI:8969253
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174612)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 174612)
Waterston,R.H.
Direct Submission
Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8748947.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0661B12
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2478: contig of 2478 bp in length
* 2479 2578: gap of unknown length
* 2579 6948: contig of 4370 bp in length
* 6949 7049: gap of unknown length
* 7049 9373: contig of 2325 bp in length
* 9374 13303: contig of 3830 bp in length
* 13304 13403: gap of unknown length
* 13404 18270: contig of 4867 bp in length
* 18271 25589: contig of 7219 bp in length
* 25590 25689: gap of unknown length
* 25690 34815: contig of 9126 bp in length
* 34816 34916: gap of unknown length
* 34916 45742: contig of 10827 bp in length
* 45743 45843: gap of unknown length
* 45843 62987: contig of 17145 bp in length
* 62988 63087: gap of unknown length
* 63088 87584: contig of 24497 bp in length
* 87585 114765: contig of 27081 bp in length
* 114766 114865: gap of unknown length
* 114866 142334: contig of 27469 bp in length
* 142335 142435: gap of unknown length
* 142435 174612: contig of 32178 bp in length.

Location/Qualifiers
1. 174612
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"

FEATURES
source

misc_feature	/clone="RP11-661B12"	
misc_feature	1. .2478	
misc_feature	/note="assembly_name:Contig4"	
misc_feature	2579. .6948	
misc_feature	/note="assembly_name:Contig5"	
misc_feature	7049. .9373	
misc_feature	/note="assembly_name:Contig6"	
misc_feature	9474. .13303	
misc_feature	/note="assembly_name:Contig7"	
misc_feature	13404. .18270	
misc_feature	/note="assembly_name:Contig8"	
misc_feature	18371. .25589	
misc_feature	/note="assembly_name:Contig9"	
misc_feature	25690. .34815	
misc_feature	/note="assembly_name:Contig10"	
misc_feature	34916. .45742	
misc_feature	/note="assembly_name:Contig11"	
misc_feature	45843. .62987	
misc_feature	/note="assembly_name:Contig12	
misc_feature	clone_end:SP6	
misc_feature	vector_side:right"	
misc_feature	63088. .87584	
misc_feature	/note="assembly_name:Contig13"	
misc_feature	87685. .114765	
misc_feature	/note="assembly_name:Contig14"	
misc_feature	114866. .142334	
misc_feature	/note="assembly_name:Contig15"	
misc_feature	142435. .174612	
misc_feature	/note="assembly_name:Contig16"	
ORIGIN		
Query Match 34.6%; Score 380.8; DB 2; Length 174612;		
Best Local Similarity 78.0%; Pred. No. 7.1e-61;		
Matches 538; Conservative 0; Mismatches 12; Indels 140; Gaps 1;		
QY	1	CACAGCAGCCACACTGTGACACTGCTTGGGCGAGCTCCCATGAGTGATGGAGGGCGGTGT 60
Db	144677	CGCAGCAGCCACACTGTGACACTGCTTGGGCGAGCTCCCATGAGTGATGGAGGGCGGTGT 144618
QY	61	CTCTGTGGCTACCGGTGAGGTGAAGGAGGGGGCCACAGGCCAGTGGCGGCTGTGCCAC 120
Db	144617	CTCTGTGGCTACCGGTGAGGTGAAGGAGGGGGCCACAGGCCAGTGGCGGCTGTGCCAC 144558
QY	121	GAGCTGGTCCCTGGACCCGAGTGTGTGTGATGTCCTGGCCCGCCGGGAGACCTACCGC 180
Db	144557	GAGCTGGTCCCTGGACCCGAGTGTGTGTGATGTCCTGGCCCGCCGGGAGACCTACCGC 144498
QY	181	TTCCGTGTGGCAGCTGTGGGCCCTGTGGTGTCTGGGGAACGGTTTCACTGCCCCGACAC 240
Db	144497	TTCCGTGTGGCAGCTGTGGGCCCTGTGGTGTCTGGGGAACGGTTTCACTGCCCCGACAC 144438
QY	241	GTGGCGGCTT----- 249
Db	144437	GTGGCGGCTTGGTGAGTTGCTTCATTGGGGTCTGGGAGTTTCAAGCCCCCACAAGCTGA 144378
QY	250	----- 249
Db	144377	GGAGGAGGGATPCCCGGGAGTGGGGCTGCACCTGGGTGATCTTTGTCTGATGGGGC 144318
QY	250	-----GAGAGCCACCGAAGCCTGTGCTTCCCGCAGC 280
Db	144317	TGTCCTCTGTCCGTGGACGCTGTTCCCGCAGCAGAGCCACGAAGCCTGTGCTTCCCGCAGC 144258
QY	281	CCTCAGCCCTGAGAGCCGGCAGTGGCAGTGGTGAAGATGCTCTCTGGAGCTTGAGG 340
Db	144257	CCTCAGCCCTGAGAGCCGGCAGTGGCAGTGGTGAAGATGCTCTCTGGAGCTTGAGG 144198
QY	341	TGGTGGCTGAGGCTGGTGAATCTTGGCACAAGGAATGGAGCGCATCCAGCCGGTG 400
Db	144197	TGGTGGCTGAGGCTGGGAGGTCTCTTGGCACAAGGAATGGAGCGCATCCAGCCGGTG 144138
QY	401	GGCGGTCGAGGTGCTTCCCGAGGTCGGCAACAGATGCTGGTGATCAAGGGCTTCACGG 460

Db	144137	GGCGGTCGAGGTGCTTCCAGGGTCGGCAACAGATGCTGGTGATCAAGGGCTTCACGG 144078		
QY	461	CAGAAGACCAAGGGGAGTACCACCTGTGGCTGGGTCCAGGGCTCCATCTGCGCTGGGGCTG 520		
Db	144077	CAGAAGACCAAGGGGAGTACCACCTGTGGCTGGGTCCAGGGCTCCATCTGCGCTGGGGCTG 144018		
QY	521	CCACCTTCCAGTGGCACTGAGCCAGCCT 550		
Db	144017	CCACCTTCCAGGTGCGTTGTCCCTGCTCT 143988		
RESULT 5				
AC026657				
LOCUS	AC026657	164766 bp DNA linear HTG 01-SEP-2000		
DEFINITION	Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT			
SEQUENCE, 31 unordered pieces.				
ACCESSION	AC026657			
VERSION	AC026657.4	GI:9958202		
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
Waterston,R.H.				
The sequence of Homo sapiens clone				
REFERENCE	2 (bases 1 to 164766)			
AUTHORS	Waterston,R.H.			
TITLE	Unpublished			
JOURNAL	2 (bases 1 to 164766)			
REFERENCE	Waterston,R.H.			
AUTHORS	Direct Submission			
TITLE	Submitted (22-MAR-2000) Genome Sequencing Center, Washington			
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
MO 63108, USA				
On Sep 1, 2000 this sequence version replaced gi:7637349.				
----- Genome Center -----				
Center: Washington University Genome Sequencing Center				
Center code: WUGSC				
Web site:http://genome.wustl.edu/gsc/index.shtml				
----- Project Information -----				
Center project name: H_NH0245P10				
----- Summary Statistics -----				
Sequencing vector: M13; 100%				
Chemistry: Dye-primer ET; 100% of reads				
Chemistry: Dye-terminator Big Dye; 0% of reads				
Assembly program: Phrap; version 0.990319				
Consensus quality: 139884 bases at least Q40				
Consensus quality: 147686 bases at least Q30				
Consensus quality: 151469 bases at least Q20				
Insert size: 169000; agarose-fp				
Insert size: 161074; sum-of-contigs				
Quality coverage: 3.60 in Q20 bases; agarose-fp				
Quality coverage: 3.92 in Q20 bases; sum-of-contigs				

* NOTE: This is a 'working draft' sequence. It currently				
* consists of 31 contigs. The true order of the pieces				
* is not known and their order in this sequence record is				
* arbitrary. Gaps between the contigs are represented as				
* runs of N, but the exact sizes of the gaps are unknown.				
* This record will be updated with the finished sequence.				
* as soon as it is available and the accession number will				
* be preserved.				

* 1 1124: contig of 1124 bp in length				
* 1125 1224: gap of unknown length				
* 1225 3032: contig of 1808 bp in length				
* 3033 3132: gap of unknown length				
* 3133 4493: contig of 1361 bp in length				
* 4494 4593: gap of unknown length				
* 4594 5860: contig of 1267 bp in length				
* 5861 5960: gap of unknown length				
* 5961 7670: contig of 1710 bp in length				
* 7671 7770: gap of unknown length				
* *				

Db	138332	CTCCGAGGCAGCCAGGAGGGTGACCTGCGATCTGCTTTGGAGGCCCTTGCTGCGAAGC	138273
Qy	641	GTGCGATGAGCGGTGAGCCACGCTGGACTCATTAGCGAGCTGCCAGAGGAGGACGCC	700
Db	138272	GTGCGATGAGCGGAGGCCACGCTGGACTCCATCAGTGAGCTGCCGAGGAAGACAGCC	138213
Qy	701	GCTCGAGCGCTCCACAGAGAGCAGAGGAGGTGGCACCTGATCTCTCTGAAGGCTACT	760
Db	138212	GTGTGAGCATCTCGGCGAGAGGAGGAGGCGGCTCTCTGACCTCTCTGAGGCTTACT	138153
Qy	761	CCAGGCGCGATGAGCTGGCCCGCACTGGAGATGTGACCTCTCACACACCACTCTGATG	820
Db	138152	CCAGGCGGATGAGCTGCGACGACAGAGAAGGTGACCTCTCACACACCACTCTGATG	138093
Qy	821	ATGAGTCCCGGGCAGGACCCCTTCCCTGGTCACTCTCAAGAAGGCTGGAGGCCAG	880
Db	138092	ACGAGTCTCGGGCTGGACCCCTTCCCTTAATTACCTACTCAGAAGGCCGGGGGTCTTG	138033
Qy	881	GCACCTCACCACTGGCCAGCAAGGTTGGGGCCCCAGCAGCC	921
Db	138032	GGATCTCACCCCTTGGCCAGCAAGGTGAGCCCTCTAGTTGGC	137992

RESULT 7	
AL645854/c	
LOCUS	211829 bp DNA linear ROD 05-APR-2002
DEFINITION	Mouse DNA sequence from clone R223-192P17 on chromosome 11, complete sequence.
ACCESSION	AL645854
VERSION	AL645854.10 GI:20068636
KEYWORDS	HTG.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

1 Smith, M.
 Direct Submission
 Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Apr 7, 2002 this sequence version replaced gi:19699553.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-192P17 is
 from the RPCI-23 Mouse PAC library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6.

FEATURES
 source
 1. 211829
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-192P17"
 /clone_lib="RPCI-23"

ORIGIN

<hr/>					
Query Match 24.3%; Score 267.6; DB 10; Length 211829;					
Best Local Similarity 80.8%; Pred. No. 5.3e-40;					
Matches 325; Conservative 0; Mismatches 74; Indels 3; Gaps 1;					
<hr/>					
Qy	524	CCTTCCAGGTGGCACTGAGCCCGACGCTCTGTGGATGAGGCC--CTCAGCCCAGCTTGCC	580		
Dd	23936	CTTCCTCAGTGGTCATGACCTCAGGCTCTGGGATGAGTCCCGCACAGCCTAGCCTGC	23877		
Qy	581	CCCCGAGGACGCCAGGAGGGTAGCTGCACCTACTCTGGGAGGCCCTTGCTCGGAAC	640		
Dd	23876	CTCCGAGGACGCCAGGAGGGTAGCTGCATCTGCTTTGGAGGCCCTTCCCGAAGC	23817		
Qy	641	GTCGCATGACCGTAGGCCACGCTGGAGTCCATTAGCGAGCTGCCAGAGGAGACGGCC	700		
Dd	23816	GCCGATGAGTCEGGAGGCCACGCTGSGATCCATCAGTAGCTGCCTGAGGAACACAGCC	23757		
Qy	701	GCTCGCAGCGCTGCCACAGGAGGACAGGAGGTGGCACCTCATCTCTGAAGGCTACT	760		
Dd	23756	GTTGTGACGATCTGGCGNAGGAGGACGAAGAAGACNGCTCCGACCTCTCTGAAGGCTACT	23697		
Qy	761	CCAGGCCGATGAGCTGGCCCGCACTCGAGATGCTGACCTCTCACACACCAGCTCTGATG	820		
Dd	23696	CTACAGCGATGAGCTAGCACGACAGGAGGCCGACCTCTCACACACCAGCTCTGACG	23637		
Qy	821	ATGAGTCCCGGACGACCCCTTCCCTGGTCACTTACCTCAAGAGGCTGGGAGGCCAG	880		
Dd	23636	ATGAGTCTCGGCTGGCACCCCTTCCCTAGTTACCTACCTCAAGAAGCCGGGGTCTCTG	23577		
Qy	881	GCACTTCACCACTGGCCACAGGTTGGGGCCCCACAGACGCC	922		
Dd	23576	GGATCTACCCCTTGGCCAGCAGGTGAGCCTCTAGTTGGCC	23535		
<hr/>					
RESULT 8					
LOCUS	QO730656	18524 bp	DNA	linear	PAT 03-FEB-2004
DEFINITION	Sequence 16590 from Patent WO02068579.				
ACCESSION	QO730656				
VERSION	QO730656.1	GI:42305092			
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1				
AUTHORS	Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.				
TITLE	Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof				
JOURNAL	Patent: WO 02068579-A 16590 06-SEP-2002;				
FEATURES	PE Corporation (NY) (US)				
source	Location/Qualifiers				
	1..18524				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
<hr/>					
ORIGIN					
<hr/>					
Query Match 22.6%; Score 248.4; DB 6; Length 18524;					
Best Local Similarity 99.6%; Pred. No. 3.4e-36;					
Matches 249; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
<hr/>					
Qy	1	CACAGAGCCACACTGTGNACACTGTCTTGGGACGCTCCCATGAGTATGGAGGCGGTGGT	60		
Dd	13422	CGCAGAGCCACACTGTGACACTGTCTTTGGGACGCTCCCATGAGTATGGAGGCGGTGGT	13481		
Qy	61	CTCTGTGGCTACCGGTGGAGTGAAGAGGGGGCCACAGGCCAGTGGCGCTGTGCCAC	120		
Dd	13482	CTCTGTGGCTACCGGTGGAGTGAAGAGGGGGCCACAGGCCAGTGGCGCTGTGCCAC	13541		
Qy	121	GAGCTGGTGCCTGGACCCGAGTGTGTGGTGGATGGCTTGGCCCCCGGGAGACCTACCGC	180		

Db 13542 GAGCTGTGCTGGAACCCGAGTGTGTGGTGTGCTGCTGCCCGGAGACCTACCGC 13601
Qy 181 TTCCGTGTGGCAGCTGTGGGCGCCCTGTGGGTGCTGGGGAACCGGTTCACTTGCCCCCAGACA 240
Db 13602 TTCCGTGTGGCAGCTGTGGGCGCCCTGTGGGTGCTGGGGAACCGGTTCACTTGCCCCCAGACA 13661
Qy 241 GTGCGGCTTG 250
Db 13662 GTGCGGCTTG 13671

RESULT 9
AC023889
LOCUS Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
DEFINITION AC023889
ACCESSION AC023889.3 GI:8969253
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 174612)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174612)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 7, 2000 this sequence version replaced gi:8748947.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0661B12
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2478: contig of 2478 bp in length
* 2479 2578: gap of unknown length
* 2579 6948: contig of 4370 bp in length
* 6949 7048: gap of unknown length
* 7049 9373: contig of 2325 bp in length
* 9374 9473: gap of unknown length
* 9474 13303: contig of 3830 bp in length
* 13304 13403: gap of unknown length
* 13404 18270: contig of 4867 bp in length
* 18271 18370: gap of unknown length

18371 25589: contig of 7219 bp in length
25590 25689: gap of unknown length
25690 34815: contig of 9126 bp in length
34816 34915: gap of unknown length
34916 45742: contig of 10827 bp in length
45743 45842: gap of unknown length
45843 62987: contig of 17145 bp in length
62988 63087: gap of unknown length
63088 87584: contig of 24497 bp in length
87585 87684: gap of unknown length
87685 114765: contig of 27081 bp in length
114766 114865: gap of unknown length
114866 142334: contig of 27469 bp in length
142335 142434: gap of unknown length
142435 174612: contig of 32178 bp in length.
FEATURES
source
1. 174612
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-661B12"
1. 2478
/note="assembly_name:Contig4"
2579. 6948
/note="assembly_name:Contig5"
7049. 9373
/note="assembly_name:Contig6"
9474. 13303
/note="assembly_name:Contig7"
13404. 18270
/note="assembly_name:Contig8"
18371. 25589
/note="assembly_name:Contig9"
25690. 34815
/note="assembly_name:Contig10"
34916. 45742
/note="assembly_name:Contig11"
45843. 62987
/note="assembly_name:Contig12"
clone_end:SP6
vector_side:right
63088. 87584
/note="assembly_name:Contig13"
87685. 114765
/note="assembly_name:Contig14"
114866. 142334
/note="assembly_name:Contig15"
142435. 174612
/note="assembly_name:Contig16"
ORIGIN
Query Match 18.2%; Score 200.8; DB 2; Length 174612;
Best Local Similarity 96.7%; Pred. No. 1.2e-27;
Matches 205; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 890 CACTGCGCAGCAAGGTGGGGCCCGCCAGCAGCCCCCTCTGTGAAGCCACAGCAGCAGCAGG 949
Db 27074 CTCGTGTCACAGGTTGGGGCCCGCCAGCAGCCCCCTCTGTGAAGCCACAGCAGCAGCAGG 27133
QY 950 AGCCACTGGCTGCTGTGGCGCCCGCCACCTGAGGAGCCTGAGCACCAGAACCTGGGTGATC 1009
Db 27134 AGCCACTGGCTGCTGTGGCGCCCGCCACCTGAGGAGCCTGAGCACCAGAACCTGGGTGATC 27193
QY 1010 CCTCAATGGACAAGGCAGCTGTGAAGATCCAGGCTGCTTTAAGGGCTTACAGGTCCCGA 1069
Db 27194 CCTCAATGGACAAGGCAGCTGTGAAGATCCAGGCTGCTTTAAGGGCTTACAGGTCCCGA 27253
QY 1070 AGGAGATGAAGCAGCAGGAAGGGGCCCATGTTTC 1101
Db 27254 AGGAGATGAAGCAGCAGGAAGGGGCCCATGTTTC 27285


```

RESULT 10
CR792456/c
LOCUS      CR792456.1       168144 bp      DNA      linear      HTG 08-OCT-2004
DEFINITION  Danio rerio clone RP71-18A8, *** SEQUENCING IN PROGRESS ***, 6
ACCESSION  CR792456
VERSION    CR792456.2  GI:54019787
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Danio rerio (zebrafish)
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
            1 (bases 1 to 168144)
REFERENCE  McLeay, K.
AUTHORS   Direct Submission
TITLE     Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL   Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Oct 9, 2004 this sequence version replaced gi:52673298.
COMMENT   ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ----- Project Information
            Center project name: bz18A8
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 166871 bases at least Q40
            Consensus quality: 166987 bases at least Q30
            Consensus quality: 167320 bases at least Q20
            Insert size: 167644; sum-of-contigs
            Insert size: 171122; 5.0% error; agarose-fp
            Quality coverage: 8.52x in Q20 bases; sum-of-contigs Quality
            coverage: 8.70x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 11201: contig of 11201 bp in length
            * 11202 11301: gap of 100 bp
            * 11302 22330: contig of 11029 bp in length
            * 22331 22430: gap of 100 bp
            * 22431 93790: contig of 71360 bp in length
            * 93791 93890: gap of 100 bp
            * 93891 106259: contig of 12369 bp in length
            * 106260 106359: gap of 100 bp
            * 106360 137391: contig of 31032 bp in length
            * 137392 137491: gap of 100 bp
            * 137492 168144: contig of 30653 bp in length.
            *
            * Location/Qualifiers
            * 1..168144
            * /organism="Danio rerio"
            * /mol_type="genomic DNA"
            * /db_xref="taxon:7955"
            * /clone_lib="RP71-18A8"
            * /clone_lib="RPCI-71"
            * 1..11201
            * /note="assembly fragment:00159
            * fragment_chain:1"
            * 11302..22330
            * /note="assembly fragment:00303
            * fragment_chain:1"
            * 22431..93790
            * /note="assembly fragment:01403
            * fragment_chain:2"

misc_feature
1..168144
/note="assembly fragment:00049
fragment_chain:2"
106360..137391
/note="assembly fragment:00928
fragment_chain:2"
137492..168144
/note="assembly fragment:00464
fragment_chain:2"
vector_side:right"

ORIGIN
Query Match 6.8%; Score 75.4; DB 2; Length 168144;
Best Local Similarity 52.9%; Pred. No. 0.00017;
Matches 282; Conservative 0; Mismatches 236; Indels 15; Gaps 5;

QY 545 CAGCCTCTGTGGATGAGGCCCTCAGCCCGCTTGGCCCCCGAGGACGCCAGGAGGGTG 604
DB 39878 CACTTTCTCAGATGCGGGCAACAGCCCAACTGCTCTCTGAGGCTGCTCTGAGGAG 39819

QY 605 ACCTGCACCTACTGTGGAGGCCCTGTGCTCGGAACGTCGCATGAGCGTCGAGCCACGC 664
DB 39818 ATTTACATGCTCTGTGGAGGCGAGGCCCAAGAGAGAAGAAATGAGTCGAGAACCCAT 39759

QY 665 TGGACTCCATTAGGAGCTGCCAGAGGAGGAGCGCGCTCGCAGGCGCTGCCACAGGAGG 724
DB 39758 TGGATTCAATTTTCAGAGTTGCCAGAAGAAGATG---GCAAAAGACAAAATGCAAAAGAAG 39702

QY 725 CAGAGGAGGTGGCACCTGATCTCTCTGAAGGCTACTCCACGCGCATGAGTGGCCCCCA 784
DB 39701 ATGAAAAGAGCCCTGCGCAAGAAATAAAGTGAGACCTGTTGAGCCAGGCCCAAAATCA 39642

QY 785 CTGAGATGCTGACCTCTTCACACACACAGCTCTGATGATGAGTCCCGGCGAGCCACCTT 844
DB 39641 CAGCAGAACCAATCTCT---ATACCAGTCTCTGATGAGAAATCTCTGAGTGGACAAAGTT 39585

QY 845 CCCTGGTCACTTACTCAAGAGGCTGGAGGCGCGAGGCACTCACCCTGCGCAGCAGG 904
DB 39584 ---TGGTGTCTTACCTTAAGAAAGACGACCAAGTC---TCCATGACAGTGGAAAGTCAA 39531

QY 905 TTGGGGCCCCAGAGCCCTCTGTGAAGCCACAGCAGCAGCAGGACCTGCGTGTGCTG 964
DB 39530 CTGAGACTGTTGCTCTTAAGAAATTTGAT---GAGCACTTCAGATGACTGAGCAATCTG 39474

QY 965 TSCGCCACCACTGGGAGACCTGAGCACCAAGACCTGGGTGATCCCTCAATGGACAAG 1024
DB 39473 TTACCAAGACATCTGTGGAGCCACTCATGGAAGCTCAGAATGGGATGAGTTGCAAGAG 39414

QY 1025 CAGCTGTGAAGATCCAGGCTGCTTTAAGGCTCAAGGTCGGAAGAGATG 1077
DB 39413 CTGCTGTCAAGATTCAGGACGCTTCAAGGGTTATAAAGCTTAGGAAGACATG 39361

RESULT 11
AF429315/c
LOCUS      AF429315.1       125020 bp      DNA      linear      PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION  AF429315
VERSION    AF429315.1  GI:17646244
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 125020)
REFERENCE  Holmes, S.B., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S.,
AUTHORS   Ingber, N.T., Ross, C.A. and Margolis, R.L.
            A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
            Nat. Genet. 29 (4), 377-378 (2001)
JOURNAL    21583737
MEDLINE

```

[illegible]

```
QY 448 AAGCGCTTCACGGCAGAGACAGACGAGGCGAGTACCACTGTGGCTGGCTCAGGCTCCATC 507
Db 198 SSCGRSGCGRRSAVSRYGTGRKGYTKMTYYSASRCWRAJYMTTYSWACSSYTWCRSK 257
QY 508 TGCCCTCGGGCTGCCACCTTCAGGTGGCACTGAGCCAGCGCTCT---GTGGATGAGGCC 564
Db 258 RRSMMWKKMRKMRWSRSYGWYSWYKMMCTAYKSYSRWCYMYRGCGWRGATRYWGRG 317
QY 565 CCTCAGCCAGCTTGCCCCCGAGCAGCAGCAGGAGGTGAGCTGACCTACTGTGGGAG 624
Db 318 YMSRMMYKMYWYRGYKGMKRWAGWAGMMWRSKRSKACYYMRWRMWRMTRRRWAK 377
QY 625 GCCCTGGCTCGGAACG---TCGCATGAGCGGTGAGCCACGCTGGACTCCATTAGCCAG 681
Db 378 KSSRTSRKKRKKMRKMRKRYKMRGYSRMRSCRRARWKRRCRSGRAWKXGRCGCMTCRMK 437
QY 682 CTGCAGAGGAGGAGCGCGCTCGCAGCGCTGCCACAGGAGGAGGAGGTGGCACCT 741
Db 438 SYGMRWKSXMRASKYKMSRMYRWKCKSRRTTWGKTRGMMGTGRCRYKRSQMK 497
QY 742 GATCTCTGAAGGCTACTCCAGCGCCGATGAGCTGGCCCGCAGCTGGAGATGCTGACCTC 801
Db 498 RKCRRRRMRMRMRKRYMSARYTMRKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 557
QY 802 TCACACACGCTCTGATGATGAGTCCCGGCGGAGCAGCAGCTTCCCTGGTACCTACCTC 861
Db 558 RYKRMYYMYKMMWYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 617
QY 862 AAGAGGCTGGGAGCGGAGCAGCTCACCACTGGCCAGCAAGGTTGGGGCCCGCAGCAGCC 921
Db 618 RKCRKASRSARAKRYAMGGHTSGSRMRWKSRYCYWRKWSKSTCTWYMYNSKYTYA 677
QY 922 CCCTCTGTGAAGCCACAGCAGCAGCAGGAGCAGCTGGCTGTGTGGCCCGCAGCAGCTGGA 981
Db 678 KYGSVRYRYAWCYMRWRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRY 737
QY 982 GACCTGACCAACAGCCTGGTGATCCCTCAATGGACA 1021
Db 738 KCSWKYRSMYVYVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWV 777
```

RESULT 13

AF429315 AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 125020)

AUTHORS Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S., Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A., Potter, N.T., Ross, C.A. and Margolis, R.L.

TITLE A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2

JOURNAL Nat. Genet. 29 (4), 377-378 (2001)

MEDLINE 21583737

PUBMED 11694876

REFERENCE 2 (bases 1 to 125020)

AUTHORS Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical

Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

FEATURES Location/Qualifiers

1..125020

source /organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="16"

repeat_region
gene
mRNA
CDS
ORIGIN
Query Match 5.3%; Score 58.4; DB 9; Length 125020;
Best Local Similarity 10.6%; Pred. No. 0.25;
Matches 103; Conservative 406; Mismatches 454; Indels 8; Gaps 1;
QY 1 CACAGCAGCCACACTGTGACACTGTCTTGGGCAGCTCCCATGATGATGGAGCGGTGT 60
Db 16998 BMSVSYVKHSHMASBSCHWBKMTWSCNMMYKSSWSSWGMCCWRRRSKGMKW 16957
QY 61 CTCTGTGCTACCGCTGAGGTGAAGAGAGGGGCCACAGCCAGTGGCGCTGTGCCAC 120
Db 16958 YSRGMSRSKRYTGGSKMMRSSMMCTSSCYASCMCCWCCSCCWCRCSCCMMR 17017
QY 121 GAGCTGGTCTGACCGGAGTGTGTGTGATGGCTGGCCCGCCGGGGAGACCTACGC 180
Db 17018 YCCACCKYMSYWTMSASYSRRSYKRSKMCNRCMRSCRSRSCSKSRGCGSGMKG 17077
QY 181 TTCCGTGTGGCAGCTGTGGGCCCTGTGGCTCTGGGAACC-----GGTTCACTGC 232
Db 17078 KSYGKTKRSRGMKAKWYMYRRMRMRKMYSSKGMYCNWGWRRRCYCSCMTSR 17137
QY 233 CCCAGACAGTGGCTTCCAGAGCCACCGAGCCTGTGCTCCCGAGCCCTCAGCCCTG 292
Db 17138 SAMCCSYCAKCKSMCYSCYYGMSMKYVYACSYRGSMSKSKYCMRGSTYSTSCG 17197
QY 293 AGACCGGAGGTGGCAGCTGTGGAAGATGTCTCTCTGGAGCTTGAGGTGGTGGTGAG 352
Db 17198 TTCCCCCNANTGGGGAAGCTTTTNCNKITYSYRKRNGCAMCKYNNYNNSSCRAGM 17257
QY 353 CTGTGTGAGTCTATCTGGCACAAGGAATGAGGCGCATCCAGCCGGTGGCGGTTCGAGG 412
Db 17258 SCTYKSMTHMSASCYWMSMYCSMRMSASRMSWSSYMMKSSWRMSYMKCCWCM 17317
QY 413 TGGTCTCCCGAGGTGGCAACAGATGTGTGTGATCAAGGGCTTCACGCGAGAAGACAGG 472
Db 17318 KMCCMRSMRSMRGSYMYVYASWKSRSRGCTCYWCMSSKSCYKSYMMMRSRKMG 17377
QY 473 GCGAGTACCACTGTGGCTGGCTCAGGCTCCATCTGCTGCGCTGCCCTGCCACCTTCCAG 532
Db 17378 KWSRCWSSWMSASRSCKYKSMRCSMMSSKCYRCAGCMWMMRGYMYRMRCSMKRW 17437
QY 533 TGGCACTGAGCCAGCTCTGTGATGAGCCCTCAGCCAGCTTGGCCCCCGAGGAG 592
Db 17438 WGSAMYMRMRWKGRCGAMCMCKYSRMSRSGKMMMSKSYCASCRSCAMSGTMYMKCASYM 17497
QY 593 CCCAGGAGGTGACCTGACACCTACTGTGGAGGAGCCCTGGCTCGGAAACGTCGCATGAGCC 652
Db 17498 CMSSYCTCWTCTSYMSYRCTCWCKGWSYTKSKSWSSMSYKGRKRSYSCMCTSRGA 17557

/map="16q24.3; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=standem
/rpt_unit="ctg"
complement(<36507..>36887)
/gene="JPH3"
/note="synonym: JP3"
complement(<36507..>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507..>36887)
/gene="JPH3"
/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"

/translation="MSSGGRPNFDDGSGCGWEDGKAHGHGVTGPKGQGEYTGSGWS
HGFEVLGVYTPSGNTYQGTWAOQRHIGLESKRWYKGEWTHGPKRGYGVRECA
NGAKYEGTWSNGLQDGYGTETYSQ"

```
QY 653 GTGAGCCACGCTGGACTCCATTAGCAGCTGCCAGAGGAGCGCGCTCGCAGCGC 712
Db 17558 MSCWRRCCYMRGASMSRRRAKGRGWSGGRSKWMTTGGMRSKYYCYTCGRMMMT 17617
QY 713 TGCACAGGAGGACGAGGAGTGCCACCTGATCTCTCTGAAGGCTACTCCACGCGCGATG 772
Db 17618 YMCCWRRRYSYMYRGSANVMGRKSSWSGMRMGASRRCKRSASRSSWCSRMRMGMRGS 17677
QY 773 AGCTGGCCCCCACTGGAGATGCTGACCTCTTCACACACCAGCTCTGATGATGATGCCGGG 832
Db 17678 CWSKMWGGRSRSASCKGSRGMRRRSKYRRGRGKRKRSKSGKSGKSKWK 17737
QY 833 CAGCACCCCTCCCTGGTCACTACTCAAGAGGCTGGGAGCGCCAGGCACTCACCAC 892
Db 17738 RSGSMTSSCYYSASSCMMWSSKSCWCCMMMKRCCACCYSSMCTSNMYRCOWGMRKSY 17797
QY 893 TGGCCAGCAGGTGGGGCCCCAGCAGCCCTCTCTGTAAGCCACAGCAGCAGGAGGAGC 952
Db 17798 STCGTCKKCTGYKSRRTWMMYSWTSKYVMWRAACMMYCYMSTRMRMMCMMSMCARKGW 17857
QY 953 CACTGGCTGCT 963
Db 17858 CMCTMGAGAT 17868

RESULT 14
AX834822
LOCUS 2534 bp DNA linear PAT 15-DEC-2003
DEFINITION Sequence 1946 from Patent EP1347046.
ACCESSION AX834822
VERSION AX834822.1 GI:39920957
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
AUTHORS
Isozaki, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahara, K. and
Masuho, Y.
TITLE Full-length cDNA sequences
JOURNAL Patent: EP 1347046-A 1946 24-SEP-2003;
RESEARCH Association for Biotechnology (JP)
FEATURES
source
1..2534
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 4.9%; Score 53.4; DB 6; Length 2534;
Best Local Similarity 53.6%; Pred. No. 5.5;
Matches 111; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 287 CCCCTGAGCGCGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 346
Db 803 CCCCTGGAAGACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 862
QY 347 CTGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 406
Db 863 GGGCGGGAACCCCGTGCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 922
QY 407 TCGAGGTGGTCTCCAGGTCGGCAACAGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 466
Db 923 ATGATGTGGTCTGGAGGCGGACGATGGCCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 982
QY 467 ACCAGGCGGAGTACCACTGGCTCGG 493
Db 983 ACGGCGGCGAGTACCACTGGTGGTGG 1009

RESULT 15
```

AK097489

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,

Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,

Sekine, M., Ohayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,

Ihii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,

Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,

Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,

Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,

Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,

Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,

Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,

Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,

Yosida, M., Hotuta, T., Kusan, J., Kanehori, K., Takahashi-Fujii, A.,

Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, P., Hara, R.,

Takeuchi, K., Arita, M., Imose, N., Musashino, K., Youki, H., Oshima, A.,

Sasaki, N., Aotsuka, S., Yoshioka, Y., Matsunawa, H., Ichihara, T.,

Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,

Teraohima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,

Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,

Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,

Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,

Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,

Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,

Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, K.,

Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsumura, K.,

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,

Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,

Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,

Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,

Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isozaki, T. and Sugano, S.

Complete sequencing and characterization of 21,243 full-length

human cDNAs

Nat. Genet. 36 (1), 40-45 (2004)

14702039

2

Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,

Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,

Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,

Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,

Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,

Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,

Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and

Isozaki, T.

NEDO human cDNA sequencing project

Unpublished

3 (bases 1 to 2534)

Isozaki, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUN-2002) Takao Isozaki, FLJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 01-438-52-3975, Fax: 01-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan

Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

HRI, and Biotechnology Center, National Institute of Technology and

Evaluation; clone selection for full insert sequencing: HRI and

RAB; annotation: HRI and RAB.

Location/Qualifiers

1..2534

/organism="Homo sapiens"

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:32:32 ; Search time 251.197 Seconds
(without alignments)
9503.485 Million cell updates/sec

Title: US-10-077-130-6_COPY_8500_8900
Perfect score: 401
Sequence: 1 gaagacaggggtggccgccccccccccaccagccagcaggagg 401

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5544816 seqs, 2976611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	100.0	2534	17	US-10-108-260A-1946
2	401	100.0	7564	18	US-10-476-397-15
3	401	100.0	14061	17	US-10-093-463-73
4	401	100.0	14109	17	US-10-093-463-71
5	401	100.0	20489	18	US-10-723-860-5701
6	401	100.0	23907	13	US-10-077-130-6
7	401	100.0	24120	13	US-10-077-130-4
8	60.6	15.1	20565	10	US-09-764-891-6045
9	60.6	15.1	20565	14	US-10-091-438-270
c 10	59.8	14.9	81940	9	US-09-759-508B-1
11	59.8	14.9	81940	10	US-09-960-706-1092

12	59.8	14.9	81940	10	US-09-873-319-724	Sequence 724, App
13	59.8	14.9	81940	18	US-10-723-860-132	Sequence 132, App
14	59.8	14.9	81940	18	US-10-656-873A-1	Sequence 1, Appli
15	59.8	14.9	82027	19	US-10-278-698-1034	Sequence 1034, Ap
16	59.8	14.9	82027	19	US-10-278-698-1045	Sequence 1045, Ap
17	56	14.0	716	10	US-09-822-846-117	Sequence 117, App
18	56	14.0	1266	10	US-09-822-846-116	Sequence 116, App
19	56	14.0	1645	9	US-09-726-643-22	Sequence 22, Appl
20	56	14.0	1645	13	US-10-042-141-22	Sequence 22, Appl
21	56	14.0	1645	19	US-10-919-272-22	Sequence 22, Appl
22	56	14.0	2170	17	US-10-264-049-850	Sequence 850, App
23	56	14.0	9591	10	US-09-764-891-6043	Sequence 6043, App
24	56	14.0	9591	14	US-10-091-438-268	Sequence 268, App
25	56	14.0	12415	10	US-09-764-891-6044	Sequence 6044, App
26	56	14.0	12415	14	US-10-051-438-269	Sequence 269, App
c 27	55.8	13.9	466	16	US-10-029-386-22631	Sequence 22631, A
28	54.4	13.6	2768	17	US-10-120-988-426	Sequence 426, App
29	47.8	11.9	3935	17	US-10-108-260A-534	Sequence 534, App
30	47.8	11.9	4200	17	US-10-341-434-140	Sequence 140, App
31	47.2	11.8	607	13	US-10-027-632-263233	Sequence 263233,
32	47.2	11.8	607	17	US-10-027-632-263233	Sequence 263233,
c 33	45.6	11.4	2254	9	US-09-726-643-44	Sequence 44, Appl
34	45.6	11.4	2254	13	US-10-042-141-44	Sequence 44, Appl
35	45.6	11.4	2254	19	US-10-919-272-44	Sequence 44, Appl
36	45.2	11.3	471	9	US-09-864-761-2416	Sequence 2416, Ap
37	45.2	11.3	573	9	US-09-864-761-14944	Sequence 14944, A
c 39	44.8	11.2	393	15	US-10-259-165-453	Sequence 453, App
c 40	44.8	11.2	396	15	US-10-259-165-121	Sequence 121, App
41	44.8	11.2	525	18	US-10-437-963-37029	Sequence 37029, A
42	44.8	11.2	2031	15	US-10-156-761-1451	Sequence 2451, Ap
c 43	44.8	11.2	9025608	15	US-10-156-761-1451	Sequence 1, Appli
44	44.4	11.1	400	17	US-10-242-535A-8373	Sequence 8373, Ap
45	44.4	11.1	400	17	US-10-085-783A-8373	Sequence 8373, Ap

ALIGNMENTS

RESULT 1
US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1946

Query Match 100.0%; Score 401; DB 17; Length 2534;
Best Local Similarity 100.0%; Pred. No. 9.7e-109;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GAAGACCAAGTGGTGGCGCCAGGGAGGACGTGGAGCTGCGCTGTGAGCTGTACGGGGC 60
Db	808	GAAGACCAAGTGGTGGCGCCAGGGAGGACGTGGAGCTGCGCTGTGAGCTGTACGGGGC 867
Qy	61	GGACGCCCGTGCACCTGGCTGAAGACAGAGGAGCCATCCGCAAGCCAGCAAGTATGAT 120
Db	868	GGACGCCCGTGCACCTGGCTGAAGACAGAGGAGCCATCCGCAAGCCAGCAAGTATGAT 927
Qy	121	GTGGTCTCGAGGGGACGATGCCATGTGCTCATCCGGGGCCCTCGCTCAAGGACGGC 180
Db	928	GTGGTCTCGAGGGGACGATGCCATGTGCTCATCCGGGGCCCTCGCTCAAGGACGGC 987

QY 181 GCGAGTACACGTGTGAGTGGAGGCTTCCAAAGACACAGCCAGGCTCCATGTGGAAGAA 240
Db |||||
988 GCGAGTACACGTGTGAGTGGAGGCTTCCAAAGACACAGCCAGGCTCCATGTGGAAGAA 1047
QY 241 AAAGCAAACTGTTTACAGAGAGAGTACCATCTGAGGTGGAGGAGAAAGGCACAGCT 300
Db |||||
1048 AAAGCAAACTGTTTACAGAGAGAGTACCATCTGAGGTGGAGGAGAAAGGCACAGCT 1107
QY 301 GTGTTTCACTGCAAGACGAGACACCCCGGGCCACAGTGACCTGGCGCAAGGGCCCTCTTG 360
Db |||||
1108 GTGTTTCACTGCAAGACGAGACACCCCGGGCCACAGTGACCTGGCGCAAGGGCCCTCTTG 1167
QY 361 GAGTACGGGCTCTAGGGAAGCACAGCCAGCCAGCCAGGAGGG 401
Db |||||
1168 GAGTACGGGCTCTAGGGAAGCACAGCCAGCCAGGAGGG 1208

RESULT 2

US-10-476-397-15
; Sequence 15, Application US/10476397
; Publication No. US20040115687A1
; GENERAL INFORMATION:

; APPLICANT: YUE, Henry
; APPLICANT: LEE, Ernestine A.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: DING, Li
; APPLICANT: JACKSON, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: LEE, Sally
; APPLICANT: WARREN, Bridget A.
; APPLICANT: XU, Yuming
; APPLICANT: TRAN, Uyen K.
; APPLICANT: LAL, Preeti G.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: YAO, Monique G.
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: GANDHI, Aneena R.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: CHINN, Anna M.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: RAMKUMAR, Javalaxmi
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: FORSYTHE, Ian J.
; TITLE OF INVENTION: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS

; FILE REFERENCE: PF-0968 USN
; CURRENT APPLICATION NUMBER: US/10/476,397
; CURRENT FILING DATE: 2003-10-31
; PRIORITY APPLICATION NUMBER: PCT/US02/13874
; PRIORITY FILING DATE: 2002-05-01
; PRIORITY APPLICATION NUMBER: US 60/288,290
; PRIORITY FILING DATE: 2001-05-02
; PRIORITY APPLICATION NUMBER: US 60/292,468
; PRIORITY FILING DATE: 2001-05-21
; PRIORITY APPLICATION NUMBER: US 60/298,616
; PRIORITY FILING DATE: 2001-06-15
; PRIORITY APPLICATION NUMBER: US 60/301,672
; PRIORITY FILING DATE: 2001-06-28
; PRIORITY APPLICATION NUMBER: US 60/345,008
; PRIORITY FILING DATE: 2002-01-04
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 7564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No: 7326129CB1
US-10-476-397-15

Query Match 100.0%; Score 401; DB 18; Length 7564;
Best Local Similarity 100.0%; Pred. No. 1.1e-108;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGACAGTGGGTGGCGCCAGGGGAGGACGTGGAGCTGGCGCTGTGAGCTTTCACGGGCG 60
Db |||||
2617 GAAGACAGTGGGTGGCGCCAGGGGAGGACGTGGAGCTGGCGCTGTGAGCTTTCACGGGCG 2676
QY 61 GGAACGCCCGGTGCACTGGCTCAAGGACAGGAAGGCCATCCCAAGAGAGCCAGAAGTATGAT 120
Db |||||
2677 GGAACGCCCGGTGCACTGGCTCAAGGACAGGAAGGCCATCCCAAGAGAGCCAGAAGTATGAT 2736
QY 121 GTGCTTCGCGAGGGCACGATGGCCATGCTGTGTCATCTCGCGGGGCGCTCGCTCAAGAGACGCG 180
Db |||||
2737 GTGCTTCGCGAGGGCACGATGGCCATGCTGTGTCATCTCGCGGGGCGCTCGCTCAAGAGACGCG 2796
QY 181 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGGCTCCATGTGGAAGAA 240
Db |||||
2797 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGGCTCCATGTGGAAGAA 2856
QY 241 AAAGCAAACTGCTTTCACAGAGGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCT 300
Db |||||
2857 AAAGCAAACTGCTTTCACAGAGGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCT 2916
QY 301 GTGTTTACAGTGCAGAGCGGAGCACCCCGCGGCCACAGTGCAGCTGGCGCAAGGGCGCTCTTG 360
Db |||||
2917 GTGTTTACAGTGCAGAGCGGAGCACCCCGCGGCCACAGTGCAGCTGGCGCAAGGGCGCTCTTG 2976
QY 361 GAGCTACGGGCTCAGGGAAGCACAGCCAGCCAGGAGGG 401
Db |||||
2977 GAGCTACGGGCTCAGGGAAGCACAGCCAGCCAGGAGGG 3017

RESULT 3

US-10-093-463-73
; Sequence 73, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Baha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypepti
; TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.


```

; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 73
; LENGTH: 14061
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(14039)
; US-10-093-463-73

```

Query Match	100.0%;	Score 401;	DB 17;	Length 14061;
Best Local Similarity	100.0%;	Pred. No. 1.2e-108;		
Matches 401;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GAAGACCACTGGGTGGCGCCAGGGAGGACGTGGAGCTGCGCTGTGAGCTGTACCGGCG	60	
DB	732	GAAGACCACTGGGTGGCGCCAGGGAGGACGTGGAGCTGCGCTGTGAGCTGTACCGGCG	791	
QY	61	GGAAAGCCCGCTGGCACTGGCTTGAAGGACAGAAAGCCATCCGCAAGACCCAGAAAGTATGAT	120	
DB	792	GGAAAGCCCGCTGGCACTGGCTTGAAGGACAGAAAGCCATCCGCAAGACCCAGAAAGTATGAT	851	
QY	121	GTGGTCTGCAGGGCACGATGGCCATGCTGGTCTATCCGCGGGGCTCGCTCAAGAGACCG	180	
DB	852	GTGGTCTGCAGGGCACGATGGCCATGCTGGTCTATCCGCGGGGCTCGCTCAAGAGACCG	911	
QY	181	GGCGAGTACAGTGTGAGTGGAGGCTTCCAAGAGCACAGCCAGCTCCCATGTGGAGAA	240	
DB	912	GGCGAGTACAGTGTGAGTGGAGGCTTCCAAGAGCACAGCCAGCTCCCATGTGGAGAA	971	
QY	241	AAAGCAAACTGCTTTCACAGAGGAGCTGACCAATCTGCAGTGGAGGAGAAAGGCACAGCT	300	
DB	972	AAAGCAAACTGCTTTCACAGAGGAGCTGACCAATCTGCAGTGGAGGAGAAAGGCACAGCT	1031	

Qy	301	GTGTTTCAGCTGCAAGACGAGACACCCCGGGCCACAGTGCACCTGGCGCAAGGCGCTCTTG	360
Db	1032	GTGTTTCAGCTGCAAGACGAGACACCCCGGGCCACAGTGCACCTGGCGCAAGGCGCTCTTG	1091
Qy	361	GAGCTACGGGCGCTCAGGGAAGACACCCAGCCAGCCAGGAGGG	401
Db	1092	GAGCTACGGGCGCTCAGGGAAGACACCCAGCCAGGAGGG	1132

RESULT 4

US-10-093-463-71

: Sequence 71, Application US/10093463

: Publication No. US20030208039A1

: GENERAL INFORMATION:

: APPLICANT: Padigaru, Muralidhara

: APPLICANT: Shenoy, Suresh

: APPLICANT: Kekuda, Ramesh

: APPLICANT: Gusev, Vladimir

: APPLICANT: Pochart, Pascal

: APPLICANT: Zhong, Mei

: APPLICANT: Rastelli, Luca

: APPLICANT: Mezes, Peter

: APPLICANT: Smithson, Glennda

: APPLICANT: Guo, Xiaojia

: APPLICANT: Gerlach, Valerie

: APPLICANT: Casman, Stacie

: APPLICANT: Boldog, Ferenc

: APPLICANT: Li, Li

: APPLICANT: Zerhusen, Bryan

: APPLICANT: Tchernev, Velizar

: APPLICANT: Gangolli, Baha

: APPLICANT: Vernet, Corine

: APPLICANT: Pena, Carol

: APPLICANT: Burgess, Catherine

: APPLICANT: Liu, Xiaohong

: APPLICANT: Spytek, Kimberly

: APPLICANT: Gorman, Linda

: APPLICANT: Spaderna, Steven

: APPLICANT: Voss, Edward

: APPLICANT: Malyankar, Uriel

: APPLICANT: Anderson, David

: APPLICANT: Patturajan, Meera

: APPLICANT: Miller, Charles

: APPLICANT: Taupier, Raymond J. Jr.

: TITLE OF INVENTION: No. US20030208039A1e1 Antibodies that Bind to Antigenic

: TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.

: FILE REFERENCE: 21402-290A (Cura 590AT)

: CURRENT APPLICATION NUMBER: US/10/093,463

: CURRENT FILING DATE: 2002-06-24

: PRIOR APPLICATION NUMBER: 60/283,675

: PRIOR FILING DATE: 2001-04-14

: PRIOR APPLICATION NUMBER: 60/338,092

: PRIOR FILING DATE: 2001-12-03

: PRIOR APPLICATION NUMBER: 60/274,281

: PRIOR FILING DATE: 2001-03-08

: PRIOR APPLICATION NUMBER: 60/274,101

: PRIOR FILING DATE: 2001-03-08

: PRIOR APPLICATION NUMBER: 60/325,681

: PRIOR FILING DATE: 2001-09-27

: PRIOR APPLICATION NUMBER: 60/304,354

: PRIOR FILING DATE: 2001-07-10

: PRIOR APPLICATION NUMBER: 60/279,995

: PRIOR FILING DATE: 2001-03-30

: PRIOR APPLICATION NUMBER: 60/294,899

: PRIOR FILING DATE: 2001-05-31

: PRIOR APPLICATION NUMBER: 60/287,424

: PRIOR FILING DATE: 2001-04-30

: PRIOR APPLICATION NUMBER: 60/299,027

: PRIOR FILING DATE: 2001-06-18

: PRIOR APPLICATION NUMBER: 60/309,198

: PRIOR FILING DATE: 2001-07-31

: PRIOR APPLICATION NUMBER: 60/281,194

```
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 14109
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(14088)
US-10-093-463-71

Query Match      100.0%; Score 401; DB 17; Length 14109;
Best Local Similarity 100.0%; Pred. No. 1.2e-108;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGACCACTGGTGGCGCCGAGGAGGACGCTGGAGCTGCCTGTGAGCTCTCACGGCG 60
Db 732 GAAGACCACTGGTGGCGCCGAGGAGGACGCTGGAGCTGCCTGTGAGCTCTCACGGCG 791

QY 61 GGAACGCCCGCTGCACTGGCTGAAGGACAGGAGGCCATCCGCAAGAGCCAGAAATGATGAT 120
Db 792 GGAACGCCCGCTGCACTGGCTGAAGGACAGGAGGCCATCCGCAAGAGCCAGAAATGATGAT 851

QY 121 GTGCTCTCGAGGACGACATGCCATGCTGTCTATCCCGGGGCTGCTCTCAAGAGCGCG 180
Db 852 GTGCTCTCGAGGACGACATGCCATGCTGTCTATCCCGGGGCTGCTCTCAAGAGCGCG 911

QY 181 GCGGAGTACACGTGTGAGGTGGAGCTTCCAAAGAGCAGCAGCCCTCCATGTGGAAGAA 240
Db 912 GCGGAGTACACGTGTGAGGTGGAGCTTCCAAAGAGCAGCAGCCCTCCATGTGGAAGAA 971

QY 241 AAAGCAAACTCTTCACAGAGAGCTGACCAATCTGCAAGTGGAGGAGAAAGGCACAGCT 300
Db 972 AAAGCAAACTCTTCACAGAGAGCTGACCAATCTGCAAGTGGAGGAGAAAGGCACAGCT 1031

QY 301 GTGTTACGTCGAAGCGGAGCACCAGCCGCGCCACAGTACCTGCGCGCAAGGGCTCTTG 360
Db 1032 GTGTTACGTCGAAGCGGAGCACCAGCCGCGCCACAGTACCTGCGCGCAAGGGCTCTTG 1091

QY 361 GAGCTACGGGCTCAGGGAAGCACCAGCCAGCCAGCAGGAGG 401
Db 1092 GAGCTACGGGCTCAGGGAAGCACCAGCCAGCCAGCAGGAGG 1132

RESULT 5
US-10-723-860-5701
; Sequence 5701, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882, 0193, NPUSQ1
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5701
; LENGTH: 20489
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (565)..(584)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2938)..(2955)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3329)..(3364)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4556)..(4586)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4721)..(4735)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4997)..(5011)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5108)..(5138)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6237)..(6300)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7520)..(7566)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12014)..(12037)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12123)..(12146)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13258)..(13277)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13918)..(13942)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15219)..(15248)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15262)..(15280)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15309)..(15381)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15652)..(15682)
; OTHER INFORMATION: n is a, c, g, or t
```

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (17286)..(17303)
; OTHER INFORMATION: n is a, c, g, or t
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (17413)..(17430)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-5701

Query Match      100.0%; Score 401; DB 18; Length 20489;
Best Local Similarity 100.0%; Pred. No. 1.2e-108;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAGACCAAGTGGTGGCCCGCCAGGAGGAGCAGTGGAGCTGCGGTGTGAGCTGTACGGGCG 60
Db 8571 GAAGACCAAGTGGTGGCCCGCCAGGAGGAGCAGTGGAGCTGCGGTGTGAGCTGTACGGGCG 8630

Qy 61 GGAACGCCCGTGCACTGGCTCAAGGACAGGAGGAGCCATCCGCAAGAGCCAGAAGTATGAT 120
Db 8631 GGAACGCCCGTGCACTGGCTCAAGGACAGGAGGAGCCATCCGCAAGAGCCAGAAGTATGAT 8690

Qy 121 GTGCTCTGCGAGGCGACGATGCCATGTGCTGCTATCCCGGGGGCTCGCTCAAGGACGCG 180
Db 8591 GTGCTCTGCGAGGCGACGATGCCATGTGCTGCTATCCCGGGGGCTCGCTCAAGGACGCG 8750

Qy 181 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCTCCATGTGGAAGAA 240
Db 8751 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCTCCATGTGGAAGAA 8810

Qy 241 AAGCAAACTCTTCCACAGAGGAGTGCACCAATCTGCAGGTGGAGGAGGAAAGGACAGCT 300
Db 8811 AAGCAAACTCTTCCACAGAGGAGTGCACCAATCTGCAGGTGGAGGAGGAAAGGACAGCT 8870

Qy 301 GTGTTTCACTGTCGAAGACGAGCAGCCCGCGCCACAGTGCCTGCGCAAGGGCTCTTG 360
Db 8871 GTGTTTCACTGTCGAAGACGAGCAGCCCGCGCCACAGTGCCTGCGCAAGGGCTCTTG 8930

Qy 361 GAGTACGGGCTCAGGGAAGCAGCCAGCCAGCCAGGAGGG 401
Db 8931 GAGTACGGGCTCAGGGAAGCAGCCAGCCAGCCAGGAGGG 8971

RESULT 6
US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-6

Query Match      100.0%; Score 401; DB 13; Length 23907;
Best Local Similarity 100.0%; Pred. No. 1.2e-108;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAGACCAAGTGGTGGCCCGCCAGGAGGAGCAGTGGAGCTGCGGTGTGAGCTGTACGGGCG 60
Db 8571 GAAGACCAAGTGGTGGCCCGCCAGGAGGAGCAGTGGAGCTGCGGTGTGAGCTGTACGGGCG 8630

Qy 61 GGAACGCCCGTGCACTGGCTCAAGGACAGGAGGAGCCATCCGCAAGAGCCAGAAGTATGAT 120
Db 8631 GGAACGCCCGTGCACTGGCTCAAGGACAGGAGGAGCCATCCGCAAGAGCCAGAAGTATGAT 8690

Qy 121 GTGCTCTGCGAGGCGACGATGCCATGTGCTGCTATCCCGGGGGCTCGCTCAAGGACGCG 180
Db 8591 GTGCTCTGCGAGGCGACGATGCCATGTGCTGCTATCCCGGGGGCTCGCTCAAGGACGCG 8750

Qy 181 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCTCCATGTGGAAGAA 240
Db 8751 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCTCCATGTGGAAGAA 8810

Qy 241 AAGCAAACTCTTCCACAGAGGAGTGCACCAATCTGCAGGTGGAGGAGGAAAGGACAGCT 300
Db 8811 AAGCAAACTCTTCCACAGAGGAGTGCACCAATCTGCAGGTGGAGGAGGAAAGGACAGCT 8870

Qy 301 GTGTTTCACTGTCGAAGACGAGCAGCCCGCGCCACAGTGCCTGCGCAAGGGCTCTTG 360
Db 8871 GTGTTTCACTGTCGAAGACGAGCAGCCCGCGCCACAGTGCCTGCGCAAGGGCTCTTG 8930

Qy 361 GAGTACGGGCTCAGGGAAGCAGCCAGCCAGCCAGGAGGG 401
Db 8931 GAGTACGGGCTCAGGGAAGCAGCCAGCCAGCCAGGAGGG 8971

RESULT 7
US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-4

Query Match      100.0%; Score 401; DB 13; Length 24120;
Best Local Similarity 100.0%; Pred. No. 1.2e-108;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAGACCAAGTGGTGGCCCGCCAGGAGGAGCAGTGGAGCTGCGGTGTGAGCTGTACGGGCG 60
Db 8571 GAAGACCAAGTGGTGGCCCGCCAGGAGGAGCAGTGGAGCTGCGGTGTGAGCTGTACGGGCG 8630

Qy 61 GGAACGCCCGTGCACTGGCTCAAGGACAGGAGGAGCCATCCGCAAGAGCCAGAAGTATGAT 120
Db 8631 GGAACGCCCGTGCACTGGCTCAAGGACAGGAGGAGCCATCCGCAAGAGCCAGAAGTATGAT 8690

Qy 121 GTGCTCTGCGAGGCGACGATGCCATGTGCTGCTATCCCGGGGGCTCGCTCAAGGACGCG 180
Db 8591 GTGCTCTGCGAGGCGACGATGCCATGTGCTGCTATCCCGGGGGCTCGCTCAAGGACGCG 8750

Qy 181 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCTCCATGTGGAAGAA 240
Db 8751 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAAGAGCACAGCCAGCTCCATGTGGAAGAA 8810
```



```
; SEQ ID NO 1
; LENGTH: 81940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (133)..(80910)
; OTHER INFORMATION:
US-09-759-508B-1

Query Match          14.9%; Score 59.8; DB 9; Length 81940;
Best Local Similarity 49.5%; Pred. No. 1.7e-07;
Matches 154; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 1 GAAGACCACTGGTGGCGCCAGGGAGGACGTGGAGCTGCGTGTGAGCTGTGCACGGCG 60
Db 17515 GAGGACCAAGCGTGAAGAGGGAGCCACTGCAGTGTGGAGTGTGAAGTCTCCAGAGAA 17574

QY 61 GGAACGCCCGTGCACCTGGCTGAAGACAGGAAGGCCATCCGCAAGAGCCAGAAAGTATGAT 120
Db 17575 AATGCTAAGGTGAATGGTTCAAAAATGGGACAGAAATCCTCAAAAGCAAGAAATATGAA 17634

QY 121 GTGGTCTGCGAGGACGATGCCATGTGTCATCCGGGGGCGCTCGCTCAAGGACGCG 180
Db 17635 AATGTTGCTGATGGCAGGGTCAGAAAACCTGTTATACATGACTGTACCCACAGAGATATT 17694

QY 181 GCGGAGTACAGCTGTGAGTGGAGCTTCCNAGAGCACAGCCAGCCTCCATGTGGAAGAA 240
Db 17695 AAAACATACACTTGTGATGTAAGGATTTTAAGACTTTCCTGTAACTGAATGCTGCGCT 17754

QY 241 AAAGCAAACTCTTCCACAGAGGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCT 300
Db 17755 CCTCATGTGGAAATCTTTAAGACCACCTACCGACCTTCAAGTTAGAGAAAAGAAATGGCT 17814

QY 301 GTGTTACAGTG 311
Db 17815 CGATTTGAGTG 17825

RESULT 12
US-09-733-319-724
; Sequence 724, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Mungier, William E.
; APPLICANT: Kulkarni, Prakash.
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 724
; LENGTH: 81940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 X90568
US-09-873-319-724

Query Match          14.9%; Score 59.8; DB 10; Length 81940;
Best Local Similarity 49.5%; Pred. No. 1.7e-07;
Matches 154; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 1 GAAGACCACTGGTGGCGCCAGGGAGGACGTGGAGCTGCGTGTGAGCTGTGCACGGCG 60
Db 17515 GAGGACCAAGCGTGAAGAGGGAGCCACTGCAGTGTGGAGTGTGAAGTCTCCAGAGAA 17574

QY 61 GGAACGCCCGTGCACCTGGCTGAAGACAGGAAGGCCATCCGCAAGAGCCAGAAAGTATGAT 120
Db 17575 AATGCTAAGGTGAATGGTTCAAAAATGGGACAGAAATCCTCAAAAGCAAGAAATATGAA 17634

QY 121 GTGGTCTGCGAGGACGATGCCATGTGTCATCCGGGGGCGCTCGCTCAAGGACGCG 180
Db 17635 AATGTTGCTGATGGCAGGGTCAGAAAACCTGTTATACATGACTGTACCCACAGAGATATT 17694

QY 181 GCGGAGTACAGCTGTGAGTGGAGCTTCCNAGAGCACAGCCAGCCTCCATGTGGAAGAA 240
Db 17695 AAAACATACACTTGTGATGTAAGGATTTTAAGACTTTCCTGTAACTGAATGCTGCGCT 17754

QY 241 AAAGCAAACTCTTCCACAGAGGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCT 300
Db 17755 CCTCATGTGGAAATCTTTAAGACCACCTACCGACCTTCAAGTTAGAGAAAAGAAATGGCT 17814

QY 301 GTGTTACAGTG 311
Db 17815 CGATTTGAGTG 17825

RESULT 11
US-09-960-706-1092
; Sequence 1092, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Mungier, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplas
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1092
; LENGTH: 81940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X90568
US-09-960-706-1092

Query Match          14.9%; Score 59.8; DB 10; Length 81940;
Best Local Similarity 49.5%; Pred. No. 1.7e-07;
Matches 154; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 1 GAAGACCACTGGTGGCGCCAGGGAGGACGTGGAGCTGCGTGTGAGCTGTGCACGGCG 60
Db 17515 GAGGACCAAGCGTGAAGAGGGAGCCACTGCAGTGTGGAGTGTGAAGTCTCCAGAGAA 17574

QY 61 GGAACGCCCGTGCACCTGGCTGAAGACAGGAAGGCCATCCGCAAGAGCCAGAAAGTATGAT 120
```

Db 17815 CGATTGAGTG 17825

RESULT 13

US-10-723-860-132
; Sequence 132, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193 NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 132
; LENGTH: 81940
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-132

Query Match 14.9%; Score 59.8; DB 18; Length 81940;
Best Local Similarity 49.5%; Pred. No. 1.7e-07;
Matches 154; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Qy 1 GAAGACCAAGTGGTGGCGCCAGGGGAGGACGTGGAGCTGCCTGTGAGCTGTACGGCG 60
Db 17515 GAGGACCAAGCGGTGCGAAGAGGGAGCCACTGCAGTGTGGAGTGTGAAGTCTCCAGAGAA 17574
Qy 61 GGAACGCCCGTGCCTGCAAGGACAGGAGGCGCATCCGCAAGAGCCAGAGATATGAT 120
Db 17575 AATGCTAAGTGAATGTCTCAAAATGGGACAGAAATCTTATACATGACTGTACCCAGAGATATT 17634
Qy 121 GTGCTCCGAGGCGACATGCCATGTGTATCCCGGGGCGCTCGCTCAAGAGCGC 180
Db 17635 AATGCTGATGATGCTGCAAGAACTTGTATACATGACTGTACCCAGAGATATT 17694
Qy 181 GCGGAGTACAGTGTGAGTGGAGCTTCCAGAGCAGCAGCGCTCCATGTGGAAGAA 240
Db 17695 AAACATACACTTGTGATGCTAAGGATTTTAAGACTTCTGTAACTGAAATGCTGCGCT 17754
Qy 241 AAAGCAAACTGCTTCACAGAGAGCTGACCAATCTGCAAGTGGAGGAGGAGGACAGCT 300
Db 17755 CCTCATGTGGAATCTTAAAGACCCTCACCAGCTTCAAGTTAGAGAGAAAGAAATGCGT 17814
Qy 301 GTGTTACGCTG 311
Db 17815 CGATTGAGTG 17825

RESULT 14

US-10-656-873A-1
; Sequence 1, Application US/10656873A
; Publication No. US20040265841A1
; GENERAL INFORMATION:
; APPLICANT: Fishman, Mark C. and Xu, Xiaolei
; TITLE OF INVENTION: Methods for Diagnosing and Treating
; TITLE OF INVENTION: Heart Disease
; FILE REFERENCE: 00786/381003
; CURRENT APPLICATION NUMBER: US/10/656,873A
; CURRENT FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: US 09/759,508
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/175,787
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1

; LENGTH: 81940
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-656-873A-1

Query Match 14.9%; Score 59.8; DB 18; Length 81940;
Best Local Similarity 49.5%; Pred. No. 1.7e-07;
Matches 154; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Qy 1 GAAGACCAAGTGGTGGCGCCAGGGGAGGACGTGGAGCTGCCTGTGAGCTGTACGGCG 60
Db 17515 GAGGACCAAGCGGTGCGAAGAGGGAGCCACTGCAGTGTGGAGTGTGAAGTCTCCAGAGAA 17574
Qy 61 GGAAGGCCCGTGCCTGCAAGGACAGGAGGCGCATCCGCAAGAGCCAGAGATATGAT 120
Db 17575 AATGCTAAGTGAATGTCTCAAAATGGGACAGAAATCTTCAAAAGCAAGAGATATGAA 17634
Qy 121 GTGCTCCGAGGCGACGATGCCATGTGTGCTATCCCGGGGCGCTCCCTCAAGGACGCG 180
Db 17635 AATGCTGCTGATGCGAGGCTCAGAAATCTTGTATACATGACTGTACCCAGAGATATT 17694
Qy 181 GCGGAGTACAGTGTGAGTGGAGCTTCCAGAGCAGCAGCGCTCCATGTGGAAGAA 240
Db 17695 AAACATACACTTGTGATGCTAAGGATTTTAAGACTTCTGTAACTGAAATGCTGCGCT 17754
Qy 241 AAAGCAAACTGCTTCACAGAGAGCTGACCAATCTGCAAGTGGAGGAGGAGGACAGCT 300
Db 17755 CCTCATGTGGAATCTTAAAGACCCTCACCAGCTTCAAGTTAGAGAGAAAGAAATGCGT 17814
Qy 301 GTGTTACGCTG 311
Db 17815 CGATTGAGTG 17825

RESULT 15

US-10-278-698-1034
; Sequence 1034, Application US/10278698
; Publication No. US2005003744A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmüller, Bruno
; APPLICANT: Haupt, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: 030027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1034
; LENGTH: 82027
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-698-1034

Query Match 14.9%; Score 59.8; DB 19; Length 82027;
Best Local Similarity 49.5%; Pred. No. 1.7e-07;
Matches 154; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Qy 1 GAAGACCAAGTGGTGGCGCCAGGGGAGGACGTGGAGCTGCCTGTGAGCTGTACGGCG 60
Db 17606 GAGGACCAAGCGGTGCGAAGAGGGAGCCACTGCAGTGTGGAGTGTGAAGTCTCCAGAGAA 17665
Qy 61 GGAAGGCCCGTGCCTGCAAGGACAGGAGGCGCATCCGCAAGAGCCAGAGATATGAT 120
Db 17666 AATGCTAAGTGAATGTCTCAAAATGGGACAGAAATCTTCAAAAGCAAGAGATATGAA 17725
Qy 121 GTGCTCCGAGGCGACGATGCCATGTGTGCTATCCCGGGGCGCTCCCTCAAGGACGCG 180
Db 17726 AATGCTGCTGATGCGAGGCTCAGAAATCTTGTATACATGACTGTACCCAGAGATATT 17785
Qy 181 GCGGAGTACAGTGTGAGTGGAGCTTCCAGAGCAGCAGCGCTCCATGTGGAAGAA 240
Db 17786 AAACATACACTTGTGATGCTAAGGATTTTAAGACTTCTGTAACTGAAATGCTGCGCT 17845

Search completed: March 21, 2005, 16:07:56
Job time : 258.197 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:32:32 ; Search time 689.694 Seconds
(without alignments)
9503.485 Million cell updates/sec

Title: US-10-077-130-6_COPY_13600_14700

Perfect score: 1101
Sequence: 1 cacacagccacactgtgac.....agcaggaaggcccatgttc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5544816 seqs, 2976611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1101	100.0	23907	13	US-10-077-130-6
2	1101	100.0	24120	13	US-10-077-130-4
3	1076	97.7	20489	18	US-10-723-860-5701
4	886.8	80.5	14061	17	US-10-093-483-73
5	886.8	80.5	14109	17	US-10-093-483-71
6	53.4	4.9	2534	17	US-10-108-260A-1946
7	53.4	4.9	7564	18	US-10-476-397-15
8	48.6	4.4	1813	18	US-10-723-860-5479
9	46.6	4.2	3613	18	US-10-723-860-4453
10	46.6	4.2	4176	18	US-10-723-860-8268
11	45.6	4.1	1529	18	US-10-437-963-48468

12	45	4.1	27705	10	US-09-942-025-12	Sequence 12, Appl
13	45	4.1	67311	10	US-09-942-025-1	Sequence 1, Appl
14	44.4	4.0	1469	18	US-10-437-963-64043	Sequence 64043, A
15	44.2	4.0	4698	15	US-10-156-761-6923	Sequence 6923, Ap
16	44.2	4.0	9025608	15	US-10-156-761-1	Sequence 1, Appl
17	43.8	4.0	815	16	US-10-257-826A-255	Sequence 255, App
18	43.8	4.0	817	16	US-10-257-826A-252	Sequence 252, App
19	43.8	4.0	819	16	US-10-257-826A-251	Sequence 251, App
20	43.8	4.0	820	16	US-10-257-826A-253	Sequence 253, App
21	43.8	4.0	825	16	US-10-257-826A-254	Sequence 254, App
22	43.8	4.0	1038	17	US-10-369-493-31577	Sequence 31577, A
23	43.8	4.0	1725	15	US-10-156-761-2800	Sequence 2800, Ap
24	43.8	4.0	9590	17	US-10-282-132A-33351	Sequence 33351, A
25	43.8	4.0	177587	13	US-10-087-192-1438	Sequence 1438, Ap
26	43.8	4.0	9025608	15	US-10-156-761-1	Sequence 1, Appl
27	43.4	3.9	1338	17	US/10/685	Sequence 35, Appl
28	43.4	3.9	1599	18	US-10-437-963-36356	Sequence 36356, A
29	43.2	3.9	1578	16	US-10-029-386-20397	Sequence 20397, A
30	43.2	3.9	3340	16	US-10-251-186-20	Sequence 20, Appl
31	43.2	3.9	3340	17	US-10-291-172-168	Sequence 168, App
32	43.2	3.9	3340	17	US-10-221-278-168	Sequence 168, App
33	43	3.9	6561	18	US-10-437-963-9359	Sequence 9359, Ap
34	42.8	3.9	2454	15	US-10-156-761-6929	Sequence 6929, Ap
35	42.6	3.9	1661	18	US-10-723-860-8081	Sequence 8081, Ap
36	42.6	3.9	2864	18	US-10-437-963-101596	Sequence 101596, A
37	42.4	3.9	474	18	US-10-425-115-88598	Sequence 88598, A
38	42.4	3.9	1192	17	US-10-260-238-3989	Sequence 3989, Ap
39	42.4	3.9	1580	18	US-10-437-963-88464	Sequence 88464, A
40	42.4	3.9	1723	17	US-10-425-114-24896	Sequence 24896, A
41	42.4	3.9	1737	17	US-10-425-114-16977	Sequence 16977, A
42	42.2	3.8	648	18	US-10-437-963-17286	Sequence 17286, A
43	42.2	3.8	1789	18	US-10-437-963-101196	Sequence 101196, A
44	42	3.8	659	16	US-10-058-597-29	Sequence 29, Appl
45	42	3.8	980	18	US-10-437-963-10642	Sequence 10642, A

ALIGNMENTS

RESULT 1

US-10-077-130-6
Sequence 6, Application US/10077130
Publication No. US20020168742A1
GENERAL INFORMATION:
APPLICANT: Kapeller-Libermann, Rosana
APPLICANT: Acton, Susan L.
TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
TITLE OF INVENTION: Members and Uses Therefor
FILE REFERENCE: MPI2001-047P/INCP1(M)
CURRENT APPLICATION NUMBER: US/10/077,130
CURRENT FILING DATE: 2002-02-15
PRIOR APPLICATION NUMBER: 60/269201
PRIOR FILING DATE: 2001-02-15
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 23907
TYPE: DNA
ORGANISM: Homo sapiens
US-10-077-130-6

Query Match 100.0%; Score 1101; DB 13; Length 23907;
Best Local Similarity 100.0%; Pred. No. 1.9e-276;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CACACAGCCACACTGTGACACTGTCTGGGGAGCTCCCATGATGATGAGCGGTGT 60
Db 13600 CACACAGCCACACTGTGACACTGTCTGGGGAGCTCCCATGATGATGAGCGGTGT 13659
Qy 61 CTCTGTGGCTACCGGTGGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGTGTGCCAC 120
Db 13660 CTCTGTGGCTACCGGTGGAGGTGAAGAGGGGGCCACAGGCCAGTGGCGGTGTGCCAC 13719


```

; LENGTH: 14061
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(14039)
US-10-093-463-73

Query Match      80.5%; Score 886.8; DB 17; Length 14061;
Best Local Similarity 98.9%; Pred. No. 9.2e-221;
Matches 904; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 1 CACAGCAGCACACGTGTGACACTGCTTGGGAGCTCCCATGAGTGTGGAGCGGTGT 60
Db ||||| 10263 AAGGTGAGCCCCC 10276
QY 9366 CGCAGCAGCACACGTGTGACACTGCTTGGGAGCTCCCATGAGTGTGGAGCGGTGT 9425
Db |||||
QY 61 CTCTGTGGCTACCGGTGTGAGGTGAAGAGGGGGCCACAGGCGCAGTGGCGGTGTGCCAC 120
Db |||||
QY 9426 CTCTGTGGCTACCGGTGTGAGGTGAAGAGGGGGCCACAGGCGCAGTGGCGGTGTGCCAC 9485
Db |||||
QY 121 GAGCTGTGTGACCGAGTGTGTGTGGATGGCTGGCCCCCGGGAGACCTTACCGC 180
Db |||||
QY 9486 GAGCTGTGTGACCGAGTGTGTGTGGATGGCTGGCCCCCGGGAGACCTTACCGC 9545
Db |||||
QY 181 TTCCGTGTGGCAGCTGTGGGCCCTGTGGTGTGGTGGGAACCGGTTCACTGCCCCAGACA 240
Db |||||
QY 9546 TTCCGTGTGGCAGCTGTGGGCCCTGTGGTGTGGTGGGAACCGGTTCACTGCCCCAGACA 9605
Db |||||
QY 241 GTCCGGCTTGACAGAGCCACCGAAGCCTGTGCCCTCCCGAGCCCTCAGCCCTCAGAGCCGG 300
Db |||||
QY 9606 GTCCGGCTT--GAGCCACCGAAGCCTGTGCCCTCCCGAGCCCTCAGCCCTCAGAGCCGG 9662
Db |||||
QY 301 CAGGTGGCAGCTGTGTGAAGATCTCTCTGGAGCTTGAGTGGTGGCTGAGGCTGGTGA 360
Db |||||
QY 9663 CAGGTGGCAGCTGTGTGAAGATCTCTCTGGAGCTTGAGTGGTGGCTGAGGCTGGTGA 9722
Db |||||
QY 361 GTCACTGTGCACAAAGGGAATGAGCGCATCCAGCCCGGTGGCGGTTGAGGTGTCTCC 420
Db |||||
QY 9723 GTCACTGTGCACAAAGGGAATGAGCGCATCCAGCCCGGTGGCGGTTGAGGTGTCTCC 9782
Db |||||
QY 421 CAGGCTGGCACAAGATGCTGTGTATCAAGGCTTTCAGGCGAAGAACAGGCGGAGTAC 480
Db |||||
QY 9783 CAGGCTGGCACAAGATGCTGTGTATCAAGGCTTTCAGGCGAAGAACAGGCGGAGTAC 9842
Db |||||
QY 481 CACTGTGGCTTGGCTCAGGGTCCATCTGCTCGCTGGCTGGCCACTTTCAGTGTGACATG 540
Db |||||
QY 9843 CACTGTGGCTTGGCTCAGGGTCCATCTGCTCGCTGGCTGGCCACTTTCAGTGTGACATG 9902
Db |||||
QY 541 AGCCAGAGCTCTGTGGATGAGCGCCCTCAGCCAGCTTGGCCCCCGAGGCGAGCCAGAG 600
Db |||||
QY 9903 AGCCAGAGCTCTGTGGATGAGCGCCCTCAGCCAGCTTGGCCCCCGAGGCGAGCCAGAG 9962
Db |||||
QY 601 GGTGACCTGCACCTACTGTGGAGGCTTGGCTCGGAAACGTGCGATGAGCGGTGAGGCC 660
Db |||||
QY 9963 GGTGACCTGCACCTACTGTGGAGGCTTGGCTCGGAAACGTGCGATGAGCGGTGAGGCC 10022
Db |||||
QY 661 AGCTGGACTCCATTAGCAGCTGCCAGAGGAGCGGCCGTTCGAGCGCTTGCACAG 720
Db |||||
QY 10023 AGCTGGACTCCATTAGCAGCTGCCAGAGGAGCGGCCGTTCGAGCGCTTGCACAG 10082
Db |||||
QY 721 GAGGCGAGGAGGTGGCAGCTGATCTCTGAAGGCTACTCAGCGCCGATGAGTGGCC 780
Db |||||
QY 10083 GAGGCGAGGAGGTGGCAGCTGATCTCTGAAGGCTACTCAGCGCCGATGAGTGGCC 10142
Db |||||
QY 781 CGACTGGAGATGCTGACCTCTCACACACAGCTCTGATGATGAGTCCCGGCGAGGAC 840
Db |||||
QY 10143 CGACTGGAGATGCTGACCTCTCACACACAGCTCTGATGATGAGTCCCGGCGAGGAC 10202
Db |||||
QY 841 CTTTCCCTGGTCACTTACCTCAAGAGGCTGGAGGCCAGGCACTTCACTTGGCCAGC 900
Db |||||
QY 10203 CTTTCCCTGGTCACTTACCTCAAGAGGCTGGAGGCCAGGCACTTCACTTGGCCAGC 10262
Db |||||
QY 901 AAGTTGGGGCCCC 914
Db |||||

RESULT 5
US-10-093-463-71
; Sequence 71, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennnda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Pastursajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: NO. US20030208039A1el Antibodies that Bind to Antigenic Polypepti
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
```

```
/ PRIOR FILING DATE: 2001-10-18
/ PRIOR APPLICATION NUMBER: 60/275,235
/ PRIOR FILING DATE: 2001-03-12
/ PRIOR APPLICATION NUMBER: 60/288,342
/ PRIOR FILING DATE: 2001-05-03
/ PRIOR APPLICATION NUMBER: 60/275,578
/ PRIOR FILING DATE: 2001-03-13
/ NUMBER OF SEQ ID NOS: 370
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 71
/ LENGTH: 14109
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (15)..(14088)
US-10-093-463-71

Query Match      80.5%; Score 886.8; DB 17; Length 14109;
Best Local Similarity 98.9%; Pred. No. 9.2e-221;
Matches 904; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY      1  CACAGCAGCCACTGTGACACTGTCTTGGCAGCTCCCATGAGTGTGAGGCGGTGTG 60
Db      9366 CGCAGCAGCCACTGTGACACTGTCTTGGCAGCTCCCATGAGTGTGAGGCGGTGTG 9425

QY      61  CTCTGTGGCTACCGCTGGAGGTGAAGAGGGGGCCACAGGCCAGTGGGGCTGTGCCAC 120
Db      9426 CTCTGTGGCTACCGCTGGAGGTGAAGAGGGGGCCACAGGCCAGTGGGGCTGTGCCAC 9485

QY      121 GAGCTGGTGGACCCGAGTGTGTGTGGATGTGGCTTGGCCCCCGGGAGACCTTACCGC 180
Db      9486 GAGCTGGTGGACCCGAGTGTGTGTGGATGTGGCTTGGCCCCCGGGAGACCTTACCGC 9545

QY      181 TTCCGTGTGCGAGCTGTGGCCCTGTGGTGTCTGGGGACCCGGTTCACCTGCCCCAGACA 240
Db      9546 TTCCGTGTGCGAGCTGTGGCCCTGTGGTGTCTGGGGAAACCGGTTTCACTGCCCCAGACA 9605

QY      241 GTCCGGCTTGCAGAGCACCCGAAGCCTGTGCCCTCCCGAGCCCTCAGCCCCCTGAGAGCCGG 300
Db      9606 GTCCGGCTT--GAGCCACCGAAGCCTGTGCCCTCCCGAGCCCTCAGCCCCCTGAGAGCCGG 9662

QY      301 CAGGTGGCAGCTGTGGAAGATGTCTCTGGAAGTGTGAGGTGGTGGCTGAGGCTGGTGAAG 360
Db      9663 CAGGTGGCAGCTGTGGAAGATGTCTCTGGAAGTGTGAGGTGGTGGCTGAGGCTGGCGAG 9722

QY      361 GTCACTCTGSCACAAGGGAATGAGCGCATCCAGCCCGGTGGCGGTTTCGAGTGTCTCC 420
Db      9723 GTCACTCTGSCACAAGGGAATGAGCGCATCCAGCCCGGTGGCGGTTTCGAGTGTCTCC 9782

QY      421 CAGGTTGCGCAACAGATGCTGGTGTATCAAGGGCTTTCAGGCGAGAACACAGGGCGAGTAC 480
Db      9783 CAGGTTGCGCAACAGATGCTGGTGTATCAAGGGCTTTCAGGCGAGAACACAGGGCGAGTAC 9842

QY      481 CACTGTGGCTGGCTCAGGGCTCCATCTGCCCCCTGGGCTGGCCACCTTCCAGGTGGCACTG 540
Db      9843 CACTGTGGCTGGCTCAGGGCTCCATCTGCCCCCTGGGCTGGCCACCTTCCAGGTGGCACTG 9902

QY      541 AGCCCGAGCTCTGTGGATGAGGCCCTCAGCCCCAGCTTGGCCCCCGAGCCAGCCAGGAG 600
Db      9903 AGCCCGAGCTCTGTGGATGAGGCCCTCAGCCCCAGCTTGGCCCCCGAGCCAGCCAGGAG 9962

QY      601 GGTGACCTGCACCTACTGTGGAGGCCCTGTGCTCGAAACGTTCGATGAGCCGTGAGCCC 660
Db      9963 GGTGACCTGCACCTACTGTGGAGGCCCTGTGCTCGAAACGTTCGATGAGCCGTGAGCCC 10022

QY      661 ACGCTGGACTCCATTAGAGCTGCCAGGAGGAGCGCCCTCGCAGCGCTGCCACAG 720
Db      10023 ACGCTGGACTCCATTAGAGCTGCCAGGAGGAGCGCCCTCGCAGCGCTGCCACAG 10082

QY      721 GAGGCAGAGGAGGTGGCACCTGATCTCTCTGAAGGCTACTCTCAGCGCCGATGAGCTGGCC 780
Db      10083 GAGGCAGAGGAGGTGGCACCTGATCTCTCTGAAGGCTACTCTCAGCGCCGATGAGCTGGCC 10142
```

```
QY      781  CGCACTGGAGATGCTGACCTCTCACACACACAGCTCTGTATGTATGATGAGTCCCGGGCAGGCACC 840
Db      10143 CGCACTGGAGATGCTGACCTCTCACACACACAGCTCTGTATGTATGATGAGTCCCGGGCAGGCACC 10202

QY      841  CCTTCCCTGGTCACTTACCTCAAGAGGCTGGGAGGCCAGGCACCTCACCACTGGCCAGC 900
Db      10203 CCTTCCCTGGTCACTTACCTCAAGAGGCTGGGAGGCCAGGCACCTCACCACTGGCCAGC 10262

QY      901  AAGTGTGGGCCCC 914
Db      10263 AAGTGTAGCCCCCC 10276

RESULT 6
US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1946

Query Match      4.9%; Score 53.4; DB 17; Length 2534;
Best Local Similarity 53.6%; Pred. No. 0.0004;
Matches 111; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY      287  CCCCTGAGAGCGCGCAGGTGGCAGCTGGTGAAGATGTCTCTCTGGAGCTTGAGGTGGTGG 346
Db      803  CCCCTGAAGACCATGAGTGGTGGCGCCAGGGAGGACGTGGAGCTGGCTGTGAGCTGTGCAC 862

QY      347  CTGAGGCTGGTGGAGTCACTCTGGCAAAAGGAATGGAGCGCATCCAGCCCGTGGCGGT 406
Db      863  GGGCGGGAACCGCCGCTGCACTGGCTGAAGGACAGGAAGGCCATCCCGCAAGAGCCAGAAAT 922

QY      407  TCAGAGTGTCTCCCAAGGTCGGCAACAGATGCTGGTGATCAAGGGCTTTCACGCGCAAG 466
Db      923  ATGATGTGGTCTGGAGGGCAGATGGCCATGCTGGTCAATCCCGGGGGCCTCGCTCAAGG 982

QY      467  ACCAGGGCGAGTACCACTGTGGCTGG 493
Db      983  ACGCGGCGAGTACACGTGTGAGGTGG 1009

RESULT 7
US-10-476-397-15
; Sequence 15, Application US/10476397
; Publication No. US20040115687A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: LEE, Ernestine A.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: DING, Li
; APPLICANT: JACKSON, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: LEE, Sally
; APPLICANT: WARREN, Bridget A.
; APPLICANT: XU, Yuming
; APPLICANT: TRAN, Uyen K.
; APPLICANT: LAL, Preeti G.
; APPLICANT: THORNTON, Michael B.
```

```
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: YAO, Monique G.
; APPLICANT: NGUYEN, Damien B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: GRIFPIN, Jennifer A.
; APPLICANT: CHINN, Anna M.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: FORSYTHE, Ian J.
; TITLE OF INVENTION: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS
; FILE REFERENCE: PF-0968 USN
; CURRENT APPLICATION NUMBER: US/10/476,397
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/US02/13874
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/288,290
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/292,468
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/298,616
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/301,672
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/345,008
; PRIOR FILING DATE: 2002-01-04
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 7564
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; FEATURE:
; OTHER INFORMATION: Incyte ID No: 7326129CB1
US-10-476-397-15

Query Match 4.9%; Score 53.4; DB 18; Length 7564;
Best Local Similarity 53.6%; Pred. No. 0.0004;
Matches 111; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 287 CCCTGAGAGCGGAGGTGGGAGCTGGTGAAGATGCTCTCTGGAGCTTGAGTGGTGG 346
DB 2612 CCCTGGAAGACAGTGGTGGGCGCCAGGAGGACGTGGAGCTGGCTGTGAGCTGTAC 2671

QY 347 CTGAGCTGTGAGGTCTCTGGCAAGGGAATGGAGCGCATCCAGCCCGTGGGCGGT 406
DB 2672 GGGCGGGAACGCCGTGCACTGGCTGAAGACAGGAGGCCATCCGAAGAGCCAGAGT 2731

QY 407 TCGAGGTGTCTCCAGGTCGGCAACAGATGCTGGTGATCAAGGGCTTTCACGGCAGAAG 466
DB 2732 ATGATGTGTCTCGGAGGCGACAGTGGCCATGCTGGTCTATCCGCGGGCTCGCTCAAG 2791

QY 467 ACCAGGCGAGTACCACTGTGGCTGG 493
DB 2792 ACGCGGCGAGTACACGTGTGAGGTGG 2818

RESULT 8
US-10-723-860-5479
; Sequence 5479, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 4453
; LENGTH: 3613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-4453

Query Match 4.2%; Score 46.6; DB 18; Length 3613;
Best Local Similarity 53.8%; Pred. No. 0.023;
Matches 119; Conservative 0; Mismatches 99; Indels 3; Gaps 1;

QY 6 CAGCCACACTGTGACACTGTCTTGGCAGCTCCCATGATGATGAGGCGGTGGTCTCTG 65
DB 1743 CAGCCAGGCGCATCACACTGACATGACATGAGCAGCACCTCGGGGCCCGCCGACATCCT 1802
QY 66 TGGCTACCGCGTGGAGGTG---AAGAGGGGGGCCACAGCCAGTGGCGGTGTGGCCACGA 122
DB 1803 GGGGTACCTGATCGAGAGCGGTAAAGGGGAGCAACACCTCGACGGCAGTGNACGCCA 1862
```

```
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 5479
; LENGTH: 1813
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (1166)..(1183)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-5479

Query Match 4.4%; Score 48.6; DB 18; Length 1813;
Best Local Similarity 52.2%; Pred. No. 0.007;
Matches 133; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

QY 4 AGCAGCCACACTGTGACACTGTCTTGGGCAGCTCCCATGATGATGAGGCGGTGGTCTC 63
DB 305 AGCAGCAGCTCTGTGACTGTGTAGCTGGGAGCCCCAGAGAGGCTGGGGAGGCTTGGCCCTC 364

QY 64 TGTGGCTTACCGGTGGAGGT---GAAGGAGGGGGGCCACAGGCGAGTGGCGGCTGTGCCAC 120
DB 365 CAGGGCTATGTCTGGAGCTCTGCAGAGAGGAGGAGCTCGAGTGGGTGCTGTGAGTGCC 424

QY 121 GAGCTGGTGGTGGACCCGAGTGTGTGTGATGGCTGGCCCCCGGGGAGAGACTTACCGC 180
DB 425 CGGCCCCATGATGTTGACCCAGCAGACTGTGCGGAACTGTGGTCTTGGGAGACAAGTTCTC 484

QY 181 TTCCGCTGTGGCAGCTGTGGGCCCTGTGGTGTCTGGGAGACCGGTTTCACTTGGCCCCAGACA 240
DB 485 CTGCGCGTGTCTGCAGTGAGTTCTGAGGGGCTGGCCCCCGGCCCATGCTTGACACGCCC 544

QY 241 GTGCGGCTTGCAGAG 255
DB 545 ATCCACATCCGAGAG 559
```

```
RESULT 9
US-10-723-860-4453
; Sequence 4453, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 4453
; LENGTH: 3613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-4453

Query Match 4.2%; Score 46.6; DB 18; Length 3613;
Best Local Similarity 53.8%; Pred. No. 0.023;
Matches 119; Conservative 0; Mismatches 99; Indels 3; Gaps 1;

QY 6 CAGCCACACTGTGACACTGTCTTGGCAGCTCCCATGATGATGAGGCGGTGGTCTCTG 65
DB 1743 CAGCCAGGCGCATCACACTGACATGACATGAGCAGCACCTCGGGGCCCGCCGACATCCT 1802
QY 66 TGGCTACCGCGTGGAGGTG---AAGAGGGGGGCCACAGCCAGTGGCGGTGTGGCCACGA 122
DB 1803 GGGGTACCTGATCGAGAGCGGTAAAGGGGAGCAACACCTCGACGGCAGTGNACGCCA 1862
```



```
; SEQ ID NO 12
; LENGTH: 27705
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-942-025-12

Query Match
Best Local Similarity 4.1%; Score 45; DB 10; Length 27705;
Matches 298; Conservative 0; Mismatches 360; Indels 7; Gaps 3;

QY 74 GCGTGGAGTGAAGGAGGGGCCACAGGCCAGTGGCGGTGTCACACAGAGCTGGTGCCTG 133
Db 26087 GCGCGGTGGCGCGCGCCAGACGACGCGTGCAGGACCTGGCGCGCGCGCTGTGGG 26146
QY 134 GACCCGAGTGTGTGTGATGCTGGTGGCCCGCGGAGACCTACCGCTTCGCTGGTGGCAG 193
Db 26147 GCGTGTTCGCGCGCGCGGAAGCGAGCACCCCGAACCGCGGCTGGCTTTGATGATGTGG 26206
QY 194 CTGTGGGCCCTGTGGGTGCTGGGAAACCGGTTACCTGCCCGCCAGACAGTGGCGTTCAG 253
Db 26207 GCACGAGCCGTGACGCTGGGTGCTGGCGGGCGCTGGCGAGCGCGCGGAGCCGG 26266
QY 254 AGCCACCGAAGCTGTGCTCCCGACCCCTCAGCCCTCAGAGCCGCGAGGTGGCAGCTG 313
Db 26267 AGCTTGGCTGCGCGGGCGCTGCGTGGCGCGCGCTGTGGCGCACAGGCGGTAG 26326
QY 314 GTGAAGATGTCTCTGTGAGCTTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 373
Db 26327 CGGAAGAGCTCACCCGAGCCCGAGCTGGACCTTGGCGGCACGGTGTGTGACCGGG 26386
QY 374 AGGGAATGAGCGCATCCAGCCCGGTGGCGGTTCGAGGTGCTTCCAGGGTTCGGCAAC 433
Db 26387 GCACAGGGGAGCTGGGTTCAGGCGGTTCGCGCGCACCTGTGTGGCGCGCACGGGGT 26446
QY 434 AGATGCTGTGTGA--TCAAGGGCTTCAAGGAGGAGTTCAGGAGGAGTTCACCTGTGGCT 491
Db 26447 ACCTTGTGTGAGTTCGCGCGCGGTGAGGCGCGCGCGCGCGCGCGCGCGCTTGTGGAT 26506
QY 492 GCGTCAAGGCTCCATCTGCGCTGGGTGCTCCAGCTTCCAGGTGGGCACTGAGCCGAGCTTC 551
Db 26507 CGCTCGCGAGTTCGCGCGCGAGAGCGGTGACGCTGGCGCGCTGCGACGTTGCGAAGC--G 26564
QY 552 TGTGATGAGGCCCTTCAGCCAGCTTGGCCCGAGGCGAGCCAGGAGGAGTTCACCTGCA 611
Db 26565 GAGAGAGTTCGCGGTGCTGGCGCGCATGACGCGCGCGCGCGCGCGCGCGCTGAGCGGTGT 26624
QY 612 CCTACTGTGGGAGGCCCTGGCTCGGAAACGTCGATGAGCCGTGAGCCCGACGCTGACTC 671
Db 26625 GCACCTGGCGCGCGCTCGACGCGCGTGTCTGCCGCGCAGACGCGCGCGCTCTC 26684
QY 672 CATTAAGCAGTTCAGAGGAGGACGCGCGCTTCGAGCGCTTCGACAGGAGGAGGAGGA 731
Db 26685 ---GCGGTGCTGGCGCGCGAAGGTGACGCGGCGCTGACCTGCACGAGCTGACGCGGGA 26741
QY 732 GGTGG 736
Db 26742 GCTGG 26746

RESULT 13
US-09-942-025-1
; Sequence 1, Application US/09942025
; Publication No. US20030054547A1
; GENERAL INFORMATION:
; APPLICANT: Kusan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENE FROM SORANGIUM
; TITLE OF INVENTION: CELLULOSUM
; FILE REFERENCE: 30062-20020.21
; CURRENT APPLICATION NUMBER: US/09/942,025
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/271,245
; PRIOR FILING DATE: 2001-02-15

; PRIOR APPLICATION NUMBER: US 09/144,085
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: US 09/010,809
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 67311
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-942-025-1

Query Match
Best Local Similarity 4.1%; Score 45; DB 10; Length 67311;
Matches 298; Conservative 0; Mismatches 360; Indels 7; Gaps 3;

QY 74 GCGTGGAGTGAAGGAGGGGCCACAGGCCAGTGGCGGTGTCACACAGAGCTGGTGCCTG 133
Db 29939 GCGCGGTGGCGCGCGCCAGACGACGCGTGCAGGACCTGGCGCGCGCGCTGTGGG 29998
QY 134 GACCCGAGTGTGTGTGATGCTGGCCCGCGGAGACCTACCGCTTCGCTGGTGGCAG 193
Db 29999 GCGTGTTCGCGCGCGCGGAAGCGAGCACCCCGAACCGCGGCTGGCTTTGATGATGTGG 30058
QY 194 CTGTGGGCCCTGTGGGTGCTGGGAAACCGGTTACCTGCCCGCCAGACAGTGGCGTTCAG 253
Db 30059 GCACGAGCCCGTGGACCTGGGCTGTGGCGCGCGCTGGCGAGCGCGGAGCCGG 30118
QY 254 AGCCACCGAAGCTGTGCTCCCGACCCCTCAGCCCTCAGAGCCGCGGAGGTTGGCAGCTG 313
Db 30119 AGCTTGGCTGCGCGGGCGCTGCGTGGCGCGCGCGCTGTGGCGCACAGGCGGTAG 30178
QY 314 GTGAAGATGTCTCTGTGAGCTTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 373
Db 30179 CGGAAGAGCTCACCCGAGCCCGAGCTGGACCTTGGCGGCACGGTGTGTGACCGGG 30238
QY 374 AGGGAATGAGCGCATCCAGCCCGGTGGCGGTTCGAGGTGCTTCCAGGGTTCGGCAAC 433
Db 30239 GCACAGGGGAGCTGGGTTCAGGCGGTTCGCGCGCACCTGTGTGGCGCGCACGGGGT 30298
QY 434 AGATGCTGTGTGA--TCAAGGGCTTCAAGGAGGAGTTCAGGAGGAGTTCACCTGTGGCT 491
Db 30299 ACCTTGTGTGAGCTTCGCGCGCGGTGGAGCGCGCGCGCGCGCGCGCTTGTGGAT 30358
QY 492 GCGTCAAGGCTCCATCTGCGCTGGGTGCTCCAGCTTCCAGGTGGGCACTGAGCCGAGCTTC 551
Db 30359 CGCTCGCGAGTTCGCGCGCGAGACGCTGACGCTGGCGCGCTGCGACGTTGCGAAGC--G 30416
QY 552 TGTGATGAGGCCCTTCAGCCAGCTTGGCCCGAGGCGAGCCAGGAGGAGTTCACCTGCA 611
Db 30417 GAGAGAGTTCGCGCTGTGTGGCGCGCATCGACGCGCGCGCGCGCTGAGCGCGGTGT 30476
QY 612 CCTACTGTGGGAGGCCCTGGCTCGGAAACGTCGATGAGCCGTGAGCCCGACGCTGACTC 671
Db 30477 GCACCTGGCGCGCGCTTCGAGCAGCGGTGTCTGCCGCGCAGACGCGCGCGCTCTC 30536
QY 672 CATTAAGCAGTTCAGAGGAGGAGCGCGCTTCGAGCGCTTCGACAGGCGCTTCAGAGGAGGAGGA 731
Db 30537 ---GCGGTGCTGGCGCGCGAAGGTGACGCGGCGCTGCACCTGCACGAGCTGACGCGGGA 30593
QY 732 GGTGG 736
Db 30594 GCTGG 30598

RESULT 14
US-10-437-963-64043
; Sequence 64043, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 16:50:48 ; Search time 1467.58 Seconds
(without alignments)
10400.615 Million cell updates/sec

Title: US-10-077-130-6_COPY_8500_8900
Perfect score: 401
Sequence: 1 gaagacacaggggtggcgcc.....caccagcccgagggagg 401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsel1:*
9: gb_gsel2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	389.4	97.1	510	5	BX478907 DKFZp686N
2	326.6	81.4	354	2	BF353892 IL5-HT070
3	97.2	24.2	742	5	BU358283 603477905
4	79.6	19.9	895	2	BE731455 601567740
5	69.2	17.3	750	2	BE734558 601570867
6	64.8	16.2	591	5	BU439885 604144795
7	64	16.0	1060	9	AL221382 Tetraodon
8	62.2	15.5	429	4	BG383506 301607 MA
9	62.2	15.5	465	2	BF602347 267554 MA
10	60.8	15.2	3789	9	AY406292 Mus muscu
11	60.6	15.1	507	4	BM729096 UI-E-E01
12	60.6	15.1	895	6	CD244577 AGENCOURT
13	60.4	15.1	765	9	BX181691 Danio rer
14	59.8	14.9	371	2	BF824977 RC3-HN002
15	59.8	14.9	723	5	BQ446463 UI-H-EU1
16	59.2	14.8	553	7	CN302349 170005326
17	59	14.7	453	9	CG544711 OST142288
18	59	14.7	1192	8	CC247513 CH261-127
19	58.6	14.6	925	9	AL053013 Drosophi
20	58.4	14.6	1074	4	BU730664 BJ730664
21	57.8	14.4	697	4	BI559709 603252581
22	57.8	14.4	801	4	BG912151 602812819
23	57.6	14.4	564	4	BG003227 MR3-GN018
24	57.4	14.3	747	7	CN039222 nm_31_f10

25	57.2	14.3	538	7	CN302346
26	57.2	14.3	588	5	BU436399
27	57	14.2	541	2	BE810116
28	56.6	14.1	505	2	BF087860
29	56.6	14.1	612	5	BU947303
30	56.4	14.1	698	7	CR787765
31	56.4	14.1	734	7	CO503162
32	56.4	14.1	782	5	BU438276
33	56.2	14.0	528	2	BE751691
34	56.2	14.0	537	4	BM666974
35	56.2	14.0	852	5	BX377135
36	56	14.0	294	4	BM695897
37	56	14.0	437	4	BM703784
38	56	14.0	516	2	BF758009
39	56	14.0	525	2	BF087734
40	56	14.0	533	2	AW995362
41	56	14.0	541	6	CA416261
42	56	14.0	565	6	CA416975
43	56	14.0	575	5	BU072543
44	56	14.0	579	1	AI818399
45	56	14.0	579	6	CA414466

ALIGNMENTS

RESULT 1
BX478907
LOCUS BX478907 510 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp686N21206_r1.686 (synonym: hlcc3) Homo sapiens cDNA clone
ACCESSION BX478907
VERSION BX478907.1 GI:31914176
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Weil, B., Amid, C., Oeanger, A.,
Fobo, G., Han, M. and Wiemann, S.
TITLE EST (Koehrer, K., Beyer, A., Mewes, H.W., Weil, B., Amid, C., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by BMFZ (Biomedical Research Center at the Heinrich-
Heine-University, Dueseldorf/Germany) within the cDNA sequencing
consortium of the German Genome Project. No sl sequence available.
This clone (DKFZp686N21206) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Location/Qualifiers
1..510
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686N21206"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Query Match 97.1%; Score 389.4; DB 5; Length 510;
Best Local Similarity 99.7%; Pred. No. 2.4e-87;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 GGGTGGCCAGGGAGGACGTGGAGCTGTGACCTGTACCGGGGAGCCCG 70

Db 6 GGTGGCGCCAGGGAGGACGTGGAGCTGCTGTGAGCTGTACCGGGCGGAACGCCCG 65
 QY 71 TGCACCTGGCTGAAGGACAGGAAGGCCATCCGAAGAGCCAGAAAGTATGATGTGCTGG 130
 Db 66 TGCACCTGGCTGAAGGACAGGAAGGCCATCCGAAGAGCCAGAAAGTATGATGTGCTGG 125
 QY 131 AGGGACAGATGCGCATGTGCTCATCCGGGGGCTCGCTCAAGGACCGGGCGAGTACA 190
 Db 126 AGGGACAGATGCGCATGTGCTCATCCGGGGGCTCGCTCAAGGACCGGGCGAGTACA 185
 QY 191 CTTGTGAGGTGAGGCTTCCAGAGCAGACGAGCTCCATGTGGAAGAAAGCAAACT 250
 Db 186 CTTGTGAGGTGAGGCTTCCAGAGCAGACGAGCTCCATGTGGAAGAAAGCAAACT 245
 QY 251 GTTTCACAGAGAGCTGACCAATCTGCAGGTGAGGAGAAAGGACAGCTGTGTTACGT 310
 Db 246 GTTTCACAGAGAGCTGACCAATCTGCAGGTGAGGAGAAAGGACAGCTGTGTTACGT 305
 QY 311 GCAAGACGAGACACCCCGGGCCACAGTACCTGCGCAAGGGCTCTTGGAGCTACGGG 370
 Db 306 GCAAGACGAGACACCCCGGGCCACAGTACCTGCGCAAGGGCTCTTGGAGCTACGGG 365
 QY 371 CCTCAGGGAAGCACCAGCCGACGAGGG 401
 Db 366 CCTCAGGGAAGCACCAGCCGACGAGGG 396

RESULT 2
 BF353892
 LOCUS IL5-HT0704-290600-108-e09 HT0704 Homo sapiens cDNA, mRNA sequence.
 BF353892
 DEFINITION EST.
 VERSION BF353892.1 GI:11312966
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 354)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
 O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 10737800
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&tl2=IL5-HT0704-
 290600-108-e09&tl3=2000-06-29&tl4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 341.

FEATURES
 source
 1. 354
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="HT0704"
 /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

ORIGIN

Query Match 81.4%; Score 326.6; DB 2; Length 354;
 Best Local Similarity 98.8%; Pred. No. 1.5e-71;
 Matches 329; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GAAGACCAAGTGGGTGGCGCCAGAGGAGCAGTGGAGCTGGCTGTGAGCTGTACAGGCG 60
 Db 17 GAAGACCAAGTGGGTGGCGCCAGAGGAGCAGTGGAGCTGGCTGTGAGCTGTACAGGCG 76
 QY 61 GGAACGCCCGTGCACCTGGCTGAAGGACAGGAAGCCATCCCAAGAGCCAGAACTATGAT 120
 Db 77 GGAACGCCCGTGCACCTGGCTGAAGGACAGGAAGCCATCCCAAGAGCCAGAACTATGAT 136
 QY 121 GTGCTCTGCGAGGCGACGATGGCCATGCTGGTCAATCCCGGGGCTCGCTCAAGGACCG 180
 Db 137 GTGCTCTGCGAGGCGACGATGGCCATGCTGGTCAATCCCGGGGCTCGCTCAAGGACCG 196
 QY 181 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAGAGCACAGCCAGCTTCCATGTGGAAGAA 240
 Db 197 GCGAGTACACGTGTGAGGTGGAGGCTTCCAAGAGCACAGCCAGCTTCCATGTGGAAGAA 256
 QY 241 AAAGCAAACTGCTTCACAGAGGAGCTGACCAATCTGAGGTGGAGGAGAAAGGCACAGCT 300
 Db 257 AAAGCAAACTGCTTCACAGAGGAGCTGACCAATCTGAGGTGGAGGAGAAAGGCACAGCT 316
 QY 301 GTGTTACAGTGCAGAGCGGAGCACCCCGCGGCC 333
 Db 317 TGTTGTCAGTGCAGAGCGGAGCACCCCGCGGCC 349

RESULT 3

BU358283/c
 LOCUS 603477905F1 CSBQCHN71 Gallus gallus cDNA clone CHEST364j2 5', mRNA
 DEFINITION sequence.
 BU358283
 ACCESSION BU358283.1 GI:25866284
 VERSION
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 742)
 Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 12445392
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1. 742
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST364j2"
 /dev_stage="36"

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

```

/lab_host="DH10B"
/clone_lib="CSFQCHN71"
/notes="Organ: hearts; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

```

ORIGIN

```

Query Match      24.2%; Score 97.2; DB 5; Length 742;
Best Local Similarity 54.9%; Pred. No. 8.7e-14;
Matches 192; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 1 GAAGACCACTGGTGGCGCCAGGAGGACGTGGAGCTGCGCTGTCAGTGTTCACGGCG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GAACACGAGGAGTGGAGGAGCGCCACCGTTATGTTGCACTGTGAGCTGACGAAACC 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 GAAACGCCGTGCACTGGCTGAAGACAGGAAAGGCCATCCGCAAGAGCCAGAAATATGAT 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 AACGACCCGCTGGAGTGGAGGAAAGAGATACGGTGTGCACTCAGGTGACAAATGATGAG 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 GTGGTCTCGAGGACGATGCCATGCTGTGTCATCCGGGGGCTCGCTCAAGACCGG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 GTCCGGCAAGAGGGACAGCTGTCAGGCTCTTCATCTACGATGCCGAGGCTCAGGATGCG 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 181 GCGAGTACACGTGTGAGTGGAGCTTCCAAAGACACAGCCAGCGCTCCATGTGAAGAA 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 GGTGATTACAGTGTGACTCAGGGGATCAACAGACCATGTCATTCGAACTCAAGTA 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 241 AAAGCAAACTGCTTCACAGAGAGCTGACCAATCTGCGAGTGGAGGAGAAAGCCAGCT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 CTACCCGCTGCTCTTTAAAGAGAGCTGAAAAATGTGGAGTCTGAAGAAGGGGAAACAGCT 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 301 GTGTTACGTGCAACAGCGGACCCCGCGGACAGTGACCTGGCGAA 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 ATCTTGCACTGCGAGATTCCAAAGCCAGATGCTCCAGTTGAGTGGA AAAA 21
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 4

```

BE731455
LOCUS      601567740F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3842552 5',
DEFINITION mRNA sequence.
ACCESSION BE731455
VERSION    BE731455.1 GI:10145447
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 895)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCW535 row: P column: 09
            High quality sequence stop: 749.

```

FEATURES

source

```

Location/Qualifiers
1..895
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3842552"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_21"
/notes="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```

ORIGIN

```

Query Match      19.9%; Score 79.6; DB 2; Length 895;
Best Local Similarity 54.0%; Pred. No. 2.4e-09;
Matches 163; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 30 CGTGAGCTGCGCTGTGAGCTGTACGGCGGGAACGCCGTGCACTGGCTGAAGGACAG 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 CGGTGTGTGAGTGTTCAGGTGTCCGGGCGCAGGCCAGGTGCGGTGGTTCAGGGCAG 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 90 GAAGGCCATCCGCAAGAGCCAGAAAGTATGTTGTCTGCGAGGGCCAGATGCCATGCT 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 TCAGGAGCTGCAGCCCGGCCCAAGTACGAGTGTGTGATGCGCTCTACCGCAAGCT 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 150 GGTCTATCCGGGGGCTCGCTCAAGGACGGGGCGAGTACAGTGTGAGTGGAGGCTTC 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 GATCATAGTATGTCTCACGAGAGGACGAGGACACCTACACCTGTGACGCGCGTGTGT 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 210 CAAGAGCACAGCAGCGCTCCATGTGGAAGAAAAGCAAACTGTTCAAGAGGAGCTGAC 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 CAAGACCACTGCACAGTCTTCGTGGAAAGAGCAATCCATCACCATTGTCGGGGGTCTGCA 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 270 CAATCTGAGTGGAGGAGAAAGGCACAGCTGTGTTTCAGTGCAGACGGAGCACCCCGC 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 GGACGTGACAGTGTGAGGCGCGCTCTGCTGCTGTTGAGTGTGAGACCTCCATCCCCCTC 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 330 GG 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 AG 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5

```

BE734558
LOCUS      601570867F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3845393 5',
DEFINITION mRNA sequence.
ACCESSION BE734558
VERSION    BE734558.1 GI:10148550
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 750)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCW543 row: f column: 18

```


REFERENCE 1
 AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fzames,C., Wincker,P., Brotier,P., Quetier,F., Saurin,W. and Weissenbach,J.
 TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
 MEDLINE 20296633
 PUBMED 10835645
 REFERENCE 2
 AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Genome Res. 10 (7), 939-949 (2000)
 MEDLINE 20359837
 PUBMED 10899143
 REFERENCE 3
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 source Location/Qualifiers
 1..1060
 /organism="Tetraodon nigroviridis"
 /mol_type="genomic DNA"
 /db_xref="taxon:99883"
 /clone="183C20"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG183BB10SP1-end : PUC-ori"

Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R., Quackenbush,J. and Keefe,J.W.
 TITLE Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly
 JOURNAL Mamm. Genome 13 (8), 475-478 (2002)
 MEDLINE 22213789
 PUBMED 12226715
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCAGGAG
 Plate: 87 row: O column: 10
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1..429
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 1P1G"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

ORIGIN

Query Match 15.5%; Score 62.2; DB 4; Length 429;
 Best Local Similarity 55.3%; Pred. No. 5 2e-05;
 Matches 121; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
 QY 10 TGGGTGGCGCCAGGGAGGACGTGGAGCTGCGCTGTGTGAGCTGTACGGCGGGAAACGCC 69
 Db 175 TGTGTGGCCCCCGGGATCCGCTGCTGAGCTGTGAACCTGCAGGCTGGCGCCCTT 234
 QY 70 GTGCACCTGGCTGAAGGACAGGAGCCATCCGCAAGAGCCAGAGTATGATGTGTCTGTC 129
 Db 235 GTCTTTTGGAGCCCAACAGGGAGGCCAGTTCAGAGGGCGAGGCGCTGGAGCTCCATGCT 294
 QY 130 GAGGCGCCCGCCCGCATCTCGTTCATCCGCGGGCTTCGCTCAAGGAGCGCGGCCGCTTAC 189
 Db 295 GAGGCGCCCGCCCGCATCTCTGTCATCCGCGCTGCAGACCCGCGCCCTCTCTAC 354
 QY 190 ACGTGTGAGTGGAGGCTTCCAAAGAGCACAGCCAGCCTC 228
 Db 355 ACCTGCCAGTCCGGGAGCGCCCGGGGCTCCAGCCTC 393

RESULT 9

BF602347
 LOCUS 267554 MARC 3BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
 DEFINITION BF602347
 ACCESSION BF602347
 VERSION BF602347.1 GI:11699571
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 465)
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.

RESULT 8

BG383506
 LOCUS 301607 MARC 1P1G Sus scrofa cDNA 5', mRNA linear EST 12-MAR-2001
 DEFINITION BG383506
 ACCESSION BG383506
 VERSION BG383506.1 GI:13307978
 KEYWORDS EST.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 429)
 AUTHORS Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)

MEDLINE 21180013

PUBMED 11282978

COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred vo.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 43 row: L column: 23
Seq primer: ATTAGGTGACACTATAG.

FEATURES

source 1..465
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

ORIGIN

Query Match 15.5%; Score 62.2; DB 2; Length 465;
Best Local Similarity 55.3%; Pred. No. 5.3e-05;
Matches 121; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 10 TGGTGGCGCCAGGGGAGACCTGAGCTGCGCTGTGAGCTGTACGGCGGGGAAACGCCCC 69
Db 83 TCGCTGGCGCCCGGGGAGCCAGTGTGCTGAGCTGTGAACTGTCCAGGCTGTGTGCCCTG 142

Qy 70 GTGCACCTGGCTGAAGCAGCAGGAAGCCATCCGACAGCCAGAGATGATGTGGTCTGC 129
Db 143 GTCTTCTGAGCCATAACGGGAAGCCGGTGAACAGCGAGGGCGTGGAGCTCCGAGCC 202

Qy 130 GAGGCGACAGATGGCCATCGTGTATCCGCGGGGCTCGCTCAAGGACGCGGGGCGAGTAC 189
Db 203 GAAGGACCCCGCGGCTCTCTGCATCGGGCTGCAGACCTGGCCCGCAGGCGCTCTAC 262

Qy 190 ACGTGTGAGTGGAGGCTTCCAGAGCACAGCCAGCCTC 228
Db 263 ACCTGCCAGTGTGGGCGAGCACCAGGGGGGCCCCAGCCTC 301

RESULT 10 AY406292 3789 bp DNA linear GSS 15-DEC-2003

LOCUS AY406292

DEFINITION Mus musculus MYBPC3 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.

ACCESSION AY406292

VERSION AY406292.1 GI:39762266

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 3789)

AUTHORS Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.

TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios

JOURNAL PUBMED REFERENCE AUTHORS

Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 3789)

Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.

TITLE

JOURNAL

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES

source 1..3789
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>3789
/gene="MYBPC3"
/locus_tag="HCM2508"

ORIGIN

Query Match 15.2%; Score 60.8; DB 9; Length 3789;
Best Local Similarity 49.4%; Pred. No. 0.00016;
Matches 158; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Qy 46 GAGCTGTACGGGCGGGAACGCCCTGCTGCTGCTGAGCAGGAGGAGCCATCCGCAAG 105
Db 1134 GAATGGCAGATCCGACGCTGAACTCAAGTGGCTTAAAGATGACAGGAGATCCAGATG 1193

Qy 106 AGCCAGAAATGATGATGCTGCTGCGAGGCGACCATGCGCATGCTGCTCATCCGCGGGCC 165
Db 1194 AGTGGCAGGTACATCTTCGAGTCCGTCGGTGCCAGGCACTGACCATCAGCCAGTGC 1253

Qy 166 TCCTCAAGACCGGGCGAGTACACGTGTGAGTGGAGCTTCCAGAGCAGCAGCCAGC 225
Db 1254 TCCTGCTGACACGACGCTACCACTGTGTGTGGGGGGGAGAAAGTGCAGACGGAG 1313

Qy 226 CTCATGTGGAAGAAAAGCAAACTGCTTCACAGAGAGCTGACCAATCTGCAGGTGGAG 285
Db 1314 CTCTTTGTCAAAGAGCCCGGTGCTGATCATCTCGGTCCCTGGAACAGCAGCTGGTATG 1373

Qy 286 GAGAAAGGCACAGCTGTGTTTCACGTGCAAGACGAGCACCCTCCGCGCCACAGTGACCTGG 345
Db 1374 GTGGGTGAGCGGTGGAGTTTGAGTGTGAGTCTCAGAGAGGGGCCCAAGTCAATGG 1433

Qy 346 CGCAAGGGCTCTTGGAGCT 365
Db 1434 CTGAAGGATGGGGTTGAGCT 1453

RESULT 11

BM729096

LOCUS BM729096 507 bp mRNA linear EST 01-MAR-2002

DEFINITION UI-E-E01-aiw-1-04-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CONTACT

UNIVERSITY

375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).
Seq primer: M13 Reverse.

FEATURES
source
1. 507
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-E01-aiw-i-04-0-UI"
/tissue_type="fetal eye"
/dev_stages="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-E01"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-E01 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 15.1%; Score 60.6; DB 4; Length 507;
Best Local Similarity 54.8%; Pred. No. 0.00013;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 TGGGTGGCGCCAGGGAGACGTGAGCTGCCTGTGAGCTGTCAAGGCGGGAACGCC 69
Db 189 TGTGTGGCCCCCGGGGAGCCAGTGTGTGAGCTGTGAACCTCTCCGGGCTGGCGCCCC 248

QY 70 GTGCACTGCTGAAGCAGGAGCCATCCGACAGGCCAGAACTATGATGTGCTCTGC 129
Db 249 GTGGTCTGAGCCACAATGGAGGCCCTGTGAGGAGGCGAGGCGCTAGAGCTCCATGCC 308

QY 130 GAGGCGACGATGGCCATGCTGTCTATCCGCGGGGCTCGCTCAAGGACGCGGGCGAGTAC 189
Db 309 GAGGCGCCCCCGCGAGTCTCTGTCATCCAGGCTGCAGGCCAGCCCATGCGAGGCTCTAC 368

QY 190 ACGTGTAGGTGGAGGCTTCCAAAGACACAGCCAGCCCTC 228
Db 369 ACCTGCCAGTCTGGAGAGCCCCCGAGGCCCAAGCCCTC 407

RESULT 12

CD244577
LOCUS CD244577 895 bp mRNA linear EST 22-MAY-2003
DEFINITION AGNCOURT 14099068 NIH MGC 180 Homo sapiens cDNA clone
IMAGE:30378371 5', mRNA sequence.
ACCESSION CD244577
VERSION CD244577.1 GI:31005041
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 895)

AUTHORS
TITLE
JOURNAL
COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: NDAM444 row: b column: 12
High quality sequence start: 22
High quality sequence stop: 631.
Location/Qualifiers

FEATURES

source

1. 895
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30378371"
/lab_host="DH10B-Ton A (T1 and T5 phage resistance)"
/clone_lib="NIH MGC 180"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 15.1%; Score 60.6; DB 6; Length 895;
Best Local Similarity 54.8%; Pred. No. 0.00015;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 10 TGGGTGGCGCCAGGGAGACGTGAGCTGCCTGTGAGCTGTCAAGGCGGGAACGCC 69
Db 60 TGTGTGGCCCCCGGGGAGCCAGTGTGTGAGCTGTGAACCTCTCCGGGCTGGCGCCCC 119

QY 70 GTGCACTGCTGAAGCAGGAGCCATCCGACAGGCCAGAACTATGATGTGCTCTGC 129
Db 120 GTGGTCTGAGCCACAATGGAGGCCCTGTGAGGAGGCGAGGCGCTAGAGCTCCATGCC 179

QY 130 GAGGCGACGATGGCCATGCTGTCTATCCGCGGGGCTCGCTCAAGGACGCGGGCGAGTAC 189
Db 180 GAGGCGCCCCCGCGAGTCTCTGTCATCCAGGCTGCAGGCCAGCCCATGCGAGGCTCTAC 239

QY 190 ACGTGTAGGTGGAGGCTTCCAAAGACACAGCCAGCCCTC 228
Db 240 ACCTGCCAGTCTGGAGAGCCCCCGAGGCCCAAGCCCTC 278

RESULT 13

CD244577
LOCUS BX181691 765 bp DNA linear GSS 13-MAR-2003
DEFINITION Danio rerio genomic clone DKEY-190D1, genomic survey sequence.
ACCESSION BX181691
VERSION BX181691.1 GI:28013507
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 765)
AUTHORS Humphray, S.J., Huckle, E. and Durham, J.L.
JOURNAL Direct Submission
Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Unpublished
This sequence was generated from the SP6 end of BAC 190D1. 190D1 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details:

http://www.sanger.ac.uk/Projects/D_rerio/.

```
FEATURES
  source
  1..765
  Location/Qualifiers
    /organism="Danio rerio"
    /mol_type="genomic DNA"
    /db_xref="taxon:7955"
    /clones="DKEY-19001"
    /issue_type="Testis"
    /note="vector pindigoBAC-536"

ORIGIN
Query Match      15.1%; Score 60.4; DB 9; Length 765;
Best Local Similarity 52.4%; Pred. No. 0.00016;
Matches 133; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 1 GAAGACAGTGGTGGCGCCAGGAGGAGCGTGGAGCTGGCTGCTGAGCTGTACGGCG 60
   |||||
Db 504 GAGGACCACTGTAGAGNAGAGCCACTGCTGAATTTGAGTGTGAAGTCTCCAAAGAG 445

QY 61 GGAACGCCCGTGCACCTGGCTGAAGGACAGGAGCCATCCGCAAGAGCCAGAGATATGAT 120
   |||||
Db 444 AAAGCAGAGTCCAGTGGTGTAGAGATGACAAAGAAATTCGCAAAACCAAGAGATATGAC 385

QY 121 GTGGTCTGGAGGACGATGGCCATGCTGGTTCATCCGGGGCGCTCGCTCAAGAGCGG 180
   |||||
Db 384 ATGGTCTGATGGCTGTAAAGAAACTCGTATTATCATGCTGCACACTTGACGACTCT 325

QY 181 GCGAGTACACGCTGTGAGGTGGAGGCTTCCAAAGACACAGCAGCGCTCCATGTGGAAGAA 240
   |||||
Db 324 AAGACGTACACCTGTGATGCTAAGCATTTCAAGACTTTCAGCTTTCAGCTTCTTAAATGTTGAGCGT 265

QY 241 AAAGCAAACCTGTT 254
   |||||
Db 264 AAGTTAACTAGATT 251

RESULT 14
BF824977/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
MEDLINE
PUBMED
COMMENT

BF824977 371 bp mRNA linear EST 13-JAN-2001
RC3-HN0028-111100-022-h09 HN0028 Homo sapiens cDNA, mRNA sequence.
BF824977
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Dias Neto,E., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.hugwig.org.br/scripts/gethtml2.pl?tl=RC3&tl2=RC3-HN0028-
111100-022-h09&tl3=2000-11-11&tl4=1)
Seq primer: puc 18 forward
High quality sequence start: 55
```


THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 20, 2005, 16:50:48 ; Search time 3663.47 Seconds
(without alignment)
10400.615 Million cell updates/sec

Title: US-10-077-130-6_COPY_2000_3000

Perfect score: 1001

Sequence: 1 cgctcaccatcgaggagtg.....ctgaggcaggggccaatgcc 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_g881:*

9: gb_g882:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	654	65.3	740	CA430292	UI-H-FL1
C 2	286.4	28.6	553	CN302349	170005326
C 3	281.6	28.1	632	CF180572	815696 MA
C 4	262	26.2	538	CN302346	170005326
C 5	255.6	25.5	522	AQ595824	HS_2132_B
C 6	243.2	24.3	621	CF179218	81316 MA
C 7	217.6	21.7	1142	BG680164	60282628
C 8	217	21.7	401	AW898791	CM0-NN007
C 9	216.4	21.6	260	BQ365446	RC6-GN007
C 10	203.4	20.3	658	AZ383233	1M0140L17
C 11	202.4	20.2	574	A1595154	ml39h10_Y
C 12	166.8	16.7	631	BG793386	UTSW_SM11
C 13	153.2	15.3	591	BH439885	604144795
C 14	149.4	14.9	343	BE159678	MR0-HT040
C 15	146.6	14.6	397	BZ837763	CH240_248
C 16	142.2	14.2	373	AA061264	ml39h10_r
C 17	136.6	13.6	786	CD099595	AGENCOURT
C 18	134.6	13.4	236	CV426951	RC6-EN008
C 19	123.4	12.3	588	BH436399	604145537
C 20	122	12.2	472	CE495069	tigr-g88-
C 21	113.8	11.4	1192	CC247513	CH261-127
C 22	98.4	9.8	655	AU119815	AU119815
C 23	98.4	9.8	724	AU141131	AU141131
C 24	96	9.6	508	CD674219	fs08a05.Y

25	94.4	9.4	770	4	BG819323	BG819323	602781512
26	93	9.3	4130	3	AK048144	Mus muscu	
27	93	9.3	5558	3	AK029863	Mus muscu	
28	92	9.2	179	4	BG989628	PM2-HT117	
29	91.2	9.1	569	4	BM254228	515592 MA	
30	88.4	8.8	2066	9	AY399870	Homo sapi	
31	88.4	8.8	2274	3	CR590353	full-leng	
32	81.4	8.1	886	5	BUS38739	AGENCOURT	
33	81.2	8.1	205	7	CV419760	RC4-CN018	
34	80.4	8.0	612	5	BU947303	1046C03.Y	
35	79.8	8.0	575	5	BU072543	1m44607.Y	
36	79.8	8.0	886	5	CR607106	full-leng	
37	79.8	8.0	886	5	BX377136	BX377136	
38	79.8	8.0	1085	4	BM546177	AGENCOURT	
C 39	78.4	7.9	980	5	BX343487	BX343487	
C 40	78.4	7.8	796	5	BX381923	BX381923	
41	77.4	7.7	2066	9	AY399872	Mus muscu	
42	76.4	7.6	1098	4	BM543727	AGENCOURT	
43	74.8	7.5	801	4	BG912151	602812819	
44	74.6	7.5	724	5	BP144357	BP144357	
C 45	74.2	7.4	533	2	AW995362	QV0-BN004	

ALIGNMENTS

RESULT 1
CA430292/2
LOCUS CA430292.1 740 bp mRNA linear EST 07-NOV-2002
DEFINITION UI-H-FL1-bf1-i-01-0-UI-s1 NCI CGAP FL1 Homo sapiens cDNA clone
UI-H-FL1-bf1-i-01-0-UI 3', mRNA sequence.
ACCESSION CA430292
VERSION CA430292.1 GI:24793018
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Cartrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 740)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: James Martin
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA-Yes.

FEATURES

Location/Qualifiers
1..740
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="UI-H-FL1-bf1-i-01-0-UI"
/tissue_type="Cell lines"
/dev_stages="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP FL1"
/note="Organ: Chondrosarcoma; Vector: pT7T3-Pac
(Pharmacia) with a modified polylinker; Site 1: Ecor I;
Site 2: Not I; NCI CGAP FL1 is a normalized cDNA library
derived from a pool of mRNA obtained from 4 cell lines
from Grade III chondrosarcoma tissues. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an Ecor I
adaptor, digested with Not I, and cloned directionally.


```

Db 358 AGAAGCTGAGCTCCAGCTCGAAAGTTTGAATGGAGGCTGTGGGCTGCACACGAGGAGCTGG 417
QY 844 TTGTGACAGCAGGAGCCAGCGGATGCCGGGAGTATAGCTGCAGAGGCTGGGGGCCAGC 903
Db 418 TGGTGACAGCAGCATGCCAGGGGACACCGGGAGTATAGCTGCCAGGGCCGGGGCCAGC 477
QY 904 GGCCTCTCTTCATCTGGATGTCAAGAGCCCAAGGTGGTGTTCGCCAAGGACC 957
Db 478 GGCCTCTCTTCAGCTGGAGCTGGCAGGTCAAGTGTCTTGTGGGCCACGAGGAGC 531

RESULT 5
AQ595824
LOCUS HS 2132 B2 A11 T7C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=2132 Col=22 Row=B, genomic survey
sequence.
ACCESSION AQ595824
VERSION AQ595824.1 GI:5021068
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
JOURNAL Kellar, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
MEDLINE Hood, L.
PUBMED Sequence-tagged connectors: A sequence approach to mapping and
COMMENT scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2132 row: B column: 22
Seq primer: T7
Class: BAC ends
High quality sequence stop: 522.
FEATURES
source
1. 522
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=2132 Col=22 Row=B"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

ORIGIN
Query Match 25.5%; Score 255.6; DB 8; Length 522;
Best Local Similarity 96.0%; Pred. No. 1e-43;
Matches 261; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 650 CTCTGAGCCCAAGTGGTGTTCCTAAGGAGCAGCTGGCAGCAGGAGGCTGCAGGAGCAGA 709
Db 251 CCCAGAGCCCAAGGTGGTGTTCCTAAGGAGCAGCTGGCAGCAGGAGGCTGCAGGAGCAGA 310
QY 710 GGACAGGACCCAGTCCCACTGAGCTGCAGGTGGCCAGGCCAGCCAGGAGGTGACGTG 769
Db 311 GGACAGGACCCAGTCCCACTGAGCTGCAGGTGGCCAGGCCAGGCCAGGAGGTGACGTG 370
QY 770 GTACAAGGATGGGAAGAGTGTGAGCTCCAGTCTCGAAAAGTGTGATGGAGGCCACAGGCTG 829
Db 371 GTACAAGGATGGGAAGAGTGTGAGCTCCAGTCTCGAAAAGTGTGATGGAGGCCACAGGCTG 430

```

```

QY 830 CACGGCAGGCTGGTTGTGCAGCAGGAGCCAGCGGATGCCGGGAGTATAGCTGCGA 889
Db 431 CACGGCANGCTGGTTGTGCATCAAGCAGGCGGATGCCGGGAGTATATCTGCTA 490
QY 890 GGCTGGGGGGCAGCGGCTCTCTTCCATCTGG 921
Db 491 GGCTGGGGGGCAGCGGCTCTCTTCCATCTGG 522

RESULT 6
CFI79218
LOCUS 813316 MARC 3PIG Sus scrofa cDNA 5', mRNA linear EST 28-JUL-2003
DEFINITION CFI79218
ACCESSION CFI79218
VERSION CFI79218.1 GI:33290994
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
TITLE Smith, T.P.L., Freking, B.A., Ford, J.J., Vallet, J.L., Fox, J.,
JOURNAL Wise, T.A., Nommaman, D.J., Wray, J.E. and Keele, J.W.
COMMENT A second set of porcine ESTs from a pooled-tissue normalized
library
Unpublished (2003)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross match v0.990329.
Plate: SRG8009 row: O column: 4
Seq primer: GTAATACGACTCACTATAGG.
FEATURES
source
1. 621
Location/Qualifiers
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3PIG"
/notes="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN
Query Match 24.3%; Score 243.2; DB 7; Length 621;
Best Local Similarity 66.7%; Pred. No. 4.4e-41;
Matches 363; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 406 TGGCGGCGCAGCTGTGTGGATGCGCTGGGGCCCGCGCAGTTTGTAGTGTGACACCT 465
Db 81 TGGCCTCGAGTGAAGTGAAGCATGGCAGGGGCCATTGCCAGCTGAGCTGAGGTGG 140
QY 466 CCGAAGCCCAAGTCCACGTGCACCTGGTACAAGGATGGCAGTGGGCGCCACTCCGGTG 525
Db 141 CCATGCCCATACGGAGGTGACTTGGTACAAGATGGGAAGAGTGTAGTTCGAGTCAA 200
QY 526 AGCGCTCTTCGAGGAGGATGTGGGACCGCGCACCGGCTGGTGCGAGCCACAGTCACCA 585
Db 201 AAGTCGCTGTGAGGCCAAGGGCTGCACCGGA---GGCTGGTGTGTCATCAAGTGGCA 257
QY 586 GGCAGGATGAAGGCACCTACTCTCGCGGTGGGGAGGACTCTGTGGACTTCCGGCTCC 645
Db 258 AGGCATATGCTGGGAGTACAGCTGGGTGCTGGGGGTCTAAGGTCTCTTCCACCTGG 317
QY 646 GCGTCTCTGAGCCCAAGGTGGTGTGTGTAAGGAGCAGCTGGCAGCAGGAGTGCAGG 705

```



```

Db      318 ATGTCACATATCCCTCAGTGTGTTTCCAGGAGCAGCCATCAGCCACTGAAGTGCAG 377
Qy      706 CAGAGGAGGAGCCAGTCCACACTGAGTGTGAGGTGCCCCAGGCCAGAGGAGTGA 765
Db      378 CCATGGCCGAGGAGGAGTCCACACTGAGTGTGAGGTGCCCCATGCCCAGACAGAGGTGA 437
Qy      766 CGTGTACAGAGTGGGAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAG 825
Db      438 CATGTGTAAGAAGTGGGAGAGCTGACTTCAAGCTCAAAAGTGGGTGTGAGGCAATG 497
Qy      826 GCTGCACCGCAGGCTGTGTGTGACGAGGAGGCCAGGCCGATGCCCGGAGATAGCT 885
Db      498 GCTGCAACCGGATGCTGTGTGTGTCATCATCGGGCAATGCATATGCTGGGAGTACAGCT 557
Qy      886 GCGAGGCTGGGGGCCAGCGGCTCTCTTCCATCTGGATGTCAAAGAGCCCAAGGTGGTGT 945
Db      558 GCGATGCTGGGGGTCATAATGCTCTCTTCCACCTGGACATCACAGAGCCCTCAGTGTGT 617
Qy      946 TTGC 949
Db      618 TTGC 621

```

```

RESULT 7
BG680164/c
LOCUS   BG680164      1142 bp      mRNA      linear      EST 01-MAY-2001
DEFINITION
602626828F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4751671 5',
mRNA sequence.
ACCESSION   BG680164
VERSION     BG680164.1  GI:13911561
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1142)
AUTHORS     NIH-MGC http://mgs.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10608 row: h column: 08
            High quality sequence stop: 708.
            Location/Qualifiers
                1..1142
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4751671"
                    /tissue_type="squamous cell carcinoma"
                    /lab_host="DH10B (T1 phage-resistant)"
                    /clone_lib="NCI CGAP Skn4"
                    /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
                    Site 2: SalI; Cloned unidirectionally. Primer: oligo dt.
                    Average insert size 1.5kb. Library constructed by Life
                    Technologies. Note: this is a NCI CGAP Library."

```

FEATURES

```

source
1..1142
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4751671"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skn4"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI CGAP Library."

```

ORIGIN

```

Query Match      21.7%; Score 217.6; DB 4; Length 1142;
Best Local Similarity 82.2%; Pred. No. 1.2e-35;
Matches 250; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
Qy      650 CTCTGAGCCCAAGGTGGTGTTCCTAAGGAGCAGCTGCGCAGCAGGAAGTGCAGGCAGA 709
Db      314 CCCAGAGCCCAAGGCGGTGTTCCTAAGGAGCAGCAGCAGGAGGTGCGCTGTA 255

```

```

Qy      710 GGCAGGAGCCAGTCCACACTGAGCTGCGAGGTGCCCCAGGCCAGAGGTGACGTG 769
Db      254 AGCAGGGGCTAGGCCCACTGAGCTGTGAGGTGCCCCAGGCCAGAGAGGTGACTTG 195
Qy      770 GTACAAGGATCGGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAGGCTG 829
Db      194 GTACAAGGACGGGAAGAAGCTGAGCTCCAGCTTGAAGTGTGATGGAGGCTGCAGGCTG 135
Qy      830 CACGCGCAGGCTGTGTGTGACGAGGAGGCCAGGCCGATGCCGGGAGTATAGTTCGA 889
Db      134 TACACGAGGCTGTGTGTGTCAGCAGGAGGCCAGGCCAGACACCCGAGAGTATAGCTGTA 75
Qy      890 GGCTGGGGGCCAGCGGCTCTCTTCCATCTCGATGTCAAAGAGCCCAAGGTGTGTTGC 949
Db      74 GGCTGGGGGCCAGCAGCTCTCTTCCGCTCAAGTGGCAGTCAAGTGTGGGAGTGC 15
Qy      950 CAAG 953
Db      14 TGAG 11

```

RESULT 8

```

AW898791
LOCUS   AW898791      401 bp      mRNA      linear      EST 24-MAY-2000
DEFINITION
CMO-NN0075-130400-332-f06 NN0075 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW898791
VERSION     AW898791.1  GI:8062996
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 401)
AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
PUBMED     10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the PAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=CMO-NN0075-130
            400-332-f06&t3=2000-04-13&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 401.
            Location/Qualifiers
                1..401
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /dev_stage="Adult"
                    /clone_lib="NN0075"
                    /note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
                    Site 2: SmaI; A mini-library was made by cloning products
                    derived from ORESTES PCR (U.S. Letters Patent application
                    No. 196,716 - Ludwig Institute for Cancer Research)
                    profiles into the pUC 18 vector. Reverse transcription of
                    tissue mRNA and cDNA amplification were performed under
                    low stringency conditions."

```

FEATURES

```

source
1..401
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN0075"
/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

ORIGIN

Query Match 21.7%; Score 217; DB 2; Length 401;
 Best Local Similarity 85.8%; Pred. No. 1.5e-35;
 Matches 241; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 650 CTCTGAGCCCAAGTGGTGTGTTGCTAAGAGCAGTGGCACGCAAGAGCTGCAGGCAGA 709
 DB 69 CCCAGAGCCCAAGGCGGTGTTGCCAAGAGCAGCAGCAGGAGGTGCAGGCTGA 128
 QY 710 GGCAGAGCCAGTGCACACTGAGTGCAGAGTGGCCAGGCCAGCAGCGAGGTGACGTG 769
 DB 129 GCGCGGGACCAAGTGCACAGCTGAGCTGCGAGGTGGCCAGGCCAGCAGAGGTGACGTG 188
 QY 770 GTACAAGAGTGGAAAGAGCTGAGCTCCAGCTCGAAAGTGTGATGGAGGCCACAGGCTG 829
 DB 189 GTACAAGAGCGGAAGAACTGAGCTCCAGCTCGAAAGTGCAGATGGAGGCCGTGGGCTG 248
 QY 830 CACGCGCAGGCTGGTGTGTCAGCAGCGCAGCGGATGCCGGGAGTATAGCTGCGA 889
 DB 249 CACACGAGGCTGGTGGTGCAGGAGCGCAGCGCAGCGCGGAGTACAGCTGCAA 308
 QY 890 GCGTGGGGCCAGCGGCTCTCTCCACTGTGATGTCAAAG 930
 DB 309 GCGCGGGGATCAGCGGCTGTCTTCCAGCTGCACGTGGCAG 349

RESULT 9

BQ365446
 LOCUS BQ365446 260 bp mRNA linear EST 21-MAY-2002
 DEFINITION RC6-GN0070-120900-022-A01 GN0070 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BQ365446
 VERSION BQ365446.1 GI:21040958
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 260)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 10737800
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&t2=RC6-GN0070-
 120900-022-A01&t3=2000-09-12&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 155
 High quality sequence stop: 228.
 Location/Qualifiers
 1. .260
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="GN0070"
 /note="Organ: placenta_normal; Vector: puc18; Site_1:

FEATURES
source

1. .260
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="GN0070"
 /note="Organ: placenta_normal; Vector: puc18; Site_1:

ORIGIN

Query Match 21.6%; Score 216.4; DB 5; Length 260;
 Best Local Similarity 95.1%; Pred. No. 1.9e-35;
 Matches 234; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 397 ACAAGAGCATGCGCGGAGCTGTGTGATGCCGTGGGGCCCGCCGCGAGTTTGAGT 456
 DB 15 ACAAGAGCATGCGCGGAGCTGTGCGATGACGGGCTGTAGGCCCGCGCAGTTTGAGT 74
 QY 457 GTGAGACTCTCGAAGCCACGTCACGTGCATCGCTGTTACAAGGATGGCATGAGCTGGGCC 516
 DB 75 GTGAGACTCTCGAAGCCACGTCACGTGCATCGCTGTTACAAGGATGGATGAGCTGACCC 134
 QY 517 AC-TCCGGTGAGCGCTTCTTTCAGGAGGATGTTGGGACGCGGACCGGCTGGTGGCAGCC 575
 DB 135 ACATCCGGAGAGGCGCTTCTTTCAGGAGGATGTTGGGACGCGGACCGGCTGGTGGCAGCC 194
 QY 576 ACAGTCAACGAGGAGGATGAAGCACCCTACTCTCTGCCGCGTGGGCGAGGACTCTGTGGAC 635
 DB 195 ACAGTCAACGAGGAGGATGAAGCACCCTACTCTCTGCCGCGTGGGCGAGGACTCTGTGGAC 254
 QY 636 TTCCGG 641
 DB 255 TTCCGG 260

RESULT 10

AZ383233
 LOCUS AZ383233 658 bp DNA linear GSS 02-OCT-2000
 DEFINITION 1M0140L17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0140L17 R, genomic survey sequence.
 ACCESSION AZ383233
 VERSION AZ383233.1 GI:10496933
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 658)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
 Niederhausern, A. and Wright, D. Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0140 row: L column: 17
 Seq primer: CACACAGGAACACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 658.
 Location/Qualifiers
 1. .658
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0140L17"

TITLE

JOURNAL

COMMENT

FEATURES
source

1. .658
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0140L17"

SmaI; Site 2: SmaI; A mini-library was made by cloning
 products derived from ORESTES PCR (U.S. Letters Patent
 application No. 196,716 - Ludwig Institute for Cancer
 Research) profiles into the pUC 18 vector. Reverse
 transcription of tissue mRNA and cDNA amplification were
 performed under low stringency conditions."

`/sex="Male"`
`/lab_host="E. Coli strain XU10-Gold, Tl-resistant, P."`
`/clone_lib="Mouse 10kb plasmid UUGC1M library"`
`/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (GI4732114 [gb|AF129072.1], a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XU10-Gold (Stratagene) cells
and selected for ampicillin resistance."`

ORIGIN

Query Match	20.3%	Score 203.4	DB 8	Length 658
Best Local Similarity	83.4%	Prod. No. 1.2e-32		
Matches 231	Conservative 0	Mismatches 46	Indels 0	Gaps 0
QY	654	GACCCCAAGTGTGTTTGGCTTAAGGAGCAGCTGGCAGCGAGGAAGCTGCAGGCAGAGGCA	713	
Db	183	GACACCAAGGCTGATGTTTGGCTCAAGGAGCAGCAGGCACGCCACCGAGGTGAAGGCTGAGCGC	242	
QY	714	GGAGCCAGTGCACACTGAGCTCGAGGTTGGCCAGGCCAGCGAGGTGACGTGGTAC	773	
Db	243	GGAAACAGTGCACACCTTGAGCTCGAGGTGGCCAGGCCAGACTGAGGTGACATGGTTC	302	
QY	774	AAGGATGGGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGCATGAGGCCACAGGCTGCACG	833	
Db	303	AAGGACGGGAGAGAGTGAAGTTCAGCTCGAAGGTGCGGTGAGAGCCTCGGGCTGCTCC	362	
QY	834	CGCAGGCTGGTTCGACGACGAGGCAGCGCGGATGCCGGGAGTATAGCTCCGAGGCT	893	
Db	363	AGGAGGCTGGTGTGACGACGAGCGGCGCAAGCGGATGCTGGGGAGTACAGCTCGGAGGCC	422	
QY	894	GGGGGCGACGGCTCTCCTTTCCATCTGGATGTCAAAG	930	
Db	423	GGGGGTGAGAAAGCTCTCTTTCGGCTGGACGTGGCAG	459	

RESULT 11
AI595154/C

AI595154	574 bp	mRNA	linear	EST 15-MAR-2000
ml39n10.y1	Straggenae mouse testis (#937308)	Mus musculus	CDNA	
clone IMAGE:514435 5', mRNA sequence.				
AI595154				
AI595154.1	GI:4604202			
EST.				
Mus musculus	(house mouse)			
Mus musculus				
Eukaryota;	Metazoa;	Chordata;	Cranialata;	Vertebrata; Euteleostomi;
Mammalia;	Eutheria;	Rodentia;	Sciurognathi;	Muridae; Murinae; Mus.
1	(bases 1 to 574)			
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Pitter,E., Swaller,T., Gibbons,M., Pope,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson.R.				
The WashU-NCI Mouse EST Project 1999				
Unpublished (1999)				
Contact: Marra M/WashU-NCI Mouse EST Project 1999				
Washington University School of Medicine				
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA				
Tel: 314 286 1900				
TITLE				
JOURNAL				
COMMENT				

Fax: 314 286 1810

Fax: 314 206 1010
Email: mouseeest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. This read is a SEQUENCE of a previously sequenced mouse clone. This read has been verified (found to hit its original self in the correct orientation)

MGI: 308283

Seq primer: -40RP from Gibco

High quality sequence stop: 418

POLYA=No.

FEATURES

source

```

1: 374
/organism="Mus musculus"

```

```
/original= pub misc  
/mol type="mrna"
```

```
/usr_type= "IKRNF"
/atrain="CD-1"
```

db ref="taxon:10090"

```
/db_xref="taxon:10090"  
/clone="IMAGE:514435"
```

```
/clone="IMAGE:514435"  
/copy="main"
```

```

"sex="males"
/sex="males"
"sex="males"
/sex="males"

```

```
/class_type="test18"
/for_c4300="10 10 week 01d"
```

```
/dev_stage="10-12 week old"
lab_host="CORP /kenneth@redhat.com"
```

```
/lab_host="SOLR (kanamycin resistant)"
```

```
/clone_lib="Stratagene mouse testis (#937308)"
```

```
/note="Organ: testis; Vector: pBluescript SK-; Site_1:
```

ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;

-5' adaptor sequence: 5' GAATTCGGCAGGAG 3'

ORIGIN

	Query Match	20.2%;	Score 202.4;	DB 1;	Length 574;
	Best Local Similarity	83.0%;	Pred. No. 1.9e-32;		
	Matches 230;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0;
Qy	654	GAGCCCAAGGTGGTGTTCCTTAAGGAGCAGCTGCACGCAGGAAGCTGCAGGCCAGAGCCA	713		
Db	464	GACACCAGGCTGATGTTTGCACGAGCAGCAGCAGCCACCGAGGTGAAGGCTGAGCGC	405		
Qy	714	GGAGCCAGTGCACACTGAGCTGCGAGGTGGCCCCAGGCCAGCGAGGTGACGTGCTAC	773		
Db	404	GGGAACAGTGTCCACCTCTAGCTGCGAGGTGGNCAGGCCCCAGACTGAGGTGACATGTTTC	345		
Qy	774	AAGGATGGGAAGAGCTGAGCTCCAGCTCGAAAGTGTGACATGGAGGCCACAGGCTGCACG	833		
Db	344	AAGGACGGGAAGAGCTGAGCTCCAGCTCGAAGTTCGCGTGTGAAGGCTCGGGCTGCTCC	285		
Qy	834	CGCAGGCTGCTGTGTGCAGCAGGACGAGCCAGCGCGATGCCGGGAGTATAGCTTCGAGGCT	893		
Db	284	AGGAGGCTGGTGGTCAGCAGCGCGGCAAGGCGGATGCTGGGGAGTACACTGCGAGGCC	225		
Qy	894	GGGGGCCAGCGGCTCTCCTTTCCATCTGGATGTCAAAG	930		
Db	224	GGGGGGCAGAAGCTCTCTCTTCGSCCTGGACGTGGCAG	188		

RESULT 12
BG793386

LOCUS	CG793386	631 bp	mRNA	linear	EST 16-MAY-2001
DEFINITION	UTSW SM11F11	UTSW Adult Mouse Skeletal cDNA clone	UTSW SM11F11, mRNA sequence.		
ACCESSION	CG793386				
VERSION	CG793386.1	GI:14128956			
KEYWORDS	EST.				
SOURCE	Mus musculus	(house mouse)			
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
AUTHORS	Gallardo, T.D., Schageman, J.J., Williams, R.S. and Shohet, R.V.				
TITLE	UT Skeletal Western Medical Center, Adult Mouse Skeletal Muscle cDNA Library				
JOURNAL	Unpublished (2001)				

Db 241 CTGCTGGCCAGAACTGACCTTCAAGATTGATGTACAGAGCCAAAACCTGTGTTTATAA 300
 QY 952 AGGACCAGGTGGCCACACAGTAGTGAGGTGAGGCTGAGGCGGGGCCCAATGCC 1001
 Db 301 ACCAGGAGAAGGTGCAGAGGGAGGTGAGTGTCTGCTGCGCAGAAAGTGCC 350

RESULT 14
 BE159678/c
 LOCUS BE159678 343 bp mRNA linear EST 21-JUN-2000
 DEFINITION MRO-HT0407-180400-015-e07 HT0407 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE159678
 VERSION BE159678.1 GI:8622399
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 343)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V.,
 O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL sequence tags

MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PUBMED 20202663

COMMENT 10737800

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml.pl?l1=4t2-MRO-HT0407-180

400-015-e07&t3=2000-04-18&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 343.

Location/Qualifiers

1..343

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stages="Adult"

/clone_lib="HT0407"

/note="Organ: head neck; Vector: puc18; Site 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

ORIGIN

Query Match 14.9%; Score 149.4; DB 2; Length 343;
 Best Local Similarity 88.7%; Pred. No. 3e-21; Indels 1; Gaps 1;
 Matches 173; Conservative 0; Mismatches 21;

QY 650 CTCTGAGCCCAAGGTGGTGTTCCTAAGGACGAGCTGGCACGACGAAAGCTGCAGGSCAGA 709

Db 228 CCCAGAGCCCAAGGTGGTGTTCCTAAGGACGAGCTGGCACGACGAAAGCTGCAGGCTGA 169

QY 710 GGCAGGAGCCCAAGTGCACACTGAGCTGGAGTGGCCCGAGGCCACAGACGAGGTGACGCTG 769

Db 168 GCGCGGGGCCAGTGCACCGCTGAGCTGCAGAGTGGCGCCCGAGGCCACAGACAGAGGTGACGCTG 109

QY 770 GTACAGAGTGGGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGATGGAGGCCACAGGCTG 829

Db 108 GTACAGAGTGGGAAGAAGCTGAGTTCAGCTCGAAAGTGCAGCTGGA-GCGTGCGCTG 50
 QY 830 CACGCGCAGGCTGGT 844
 Db 49 CACACGGAGGCTGGT 35

RESULT 15
 BZ837763
 LOCUS BZ837763 397 bp DNA linear GSS 18-MAR-2003
 DEFINITION CH240_248L16_TV CHORI-240 Bos taurus genomic clone CH240_248L16,
 genomic survey sequence.

ACCESSION BZ837763
 VERSION BZ837763.1 GI:29065122
 KEYWORDS GSS.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 REFERENCE 1 (bases 1 to 397)
 AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
 Shvartsbeyn,A., Gebregorgis,E., Chen,D., Riggs,P., de Jong,P.,
 Crawford,A.M. and McEwan,J.C.

Bovine BAC End Sequences from Library CHORI-240
 Unpublished (2003)
 JOURNAL Contact: Shaying Zhao
 COMMENT Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the bovine BAC library CHORI-240
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/ordering_information.html). This work
 was undertaken as part of the International Bovine BAC Mapping
 Consortium (IBBMC) by AgResearch Ltd., New Zealand and The
 Institute of Genomic Research (TIGR), USA.

Plate: 248 row: L column: 16

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..397

/organism="Bos taurus"

/mol_type="genomic DNA"

/strain="breed: Hereford"

/db_xref="taxon:9913"

/clones="CH240_248L16"

/sex="Male"

/cell_type="Blood"

/clone_lib="CHORI-240"

/note="Vector: pFARBAC1.3; Site 1: MboI; Site 2: MboI;
 Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
 library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 14.6%; Score 146.6; DB 8; Length 397;
 Best Local Similarity 85.0%; Pred. No. 1.2e-20;
 Matches 164; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 650 CTCTGAGCCCAAGGTGGTGTTCCTAAGGACGAGCTGGCACGACGAAAGCTGCAGGSCAGA 709

Db 205 CCCAGAGCCCTCGGTGGTGTTCGCAAGGACGAGCGCCGACGAGTGAAGTGCAGGCCGT 264

QY 710 GGCAGGAGCCCAAGTGCACACTGAGCTGGAGTGGCCCGAGGCCACAGACGAGGTGACGCTG 769

Db 265 GCGCGGAACCAAGTGCACACTGAGCTGCAGAGTGGCGCCCGAGGCCACAGACGAGGTGATGTG 324

QY 770 GTACAGAGTGGGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGATGGAGGCCACAGGCTG 829

Db 325 GTACAGGAGCGCAAGAGCTGAGTTGAGCTCGAAAGTCGTGTGGAGGCCACGGCGC 384

Qy 830 CACGCGCAGGCTG 842

Db 385 CGGGCGGGCGCTG 397

Search completed: March 21, 2005, 14:13:54
Job time : 3665.47 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 16:50:48 ; Search time 10983.1 Seconds
(without alignments)
10400.615 Million cell updates/sec

Title: US-10-077-130-6_COPY_19500_22500

Perfect score: 3001

Sequence: 1 cctacggagcagaactcggg.....ccacctggagcaaacagcga 3001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

1: gb_est1.*

2: gb_est2.*

3: gb_hic.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_gest1.*

9: gb_gest2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	867	28.9	907	5	BX460671
2	541	18.0	622	4	BG999843
3	482.2	16.1	723	7	CK357545
4	434	14.5	517	4	BG989614
5	307	10.2	920	2	BF160452
6	281.6	9.4	429	1	AI604658
7	279	9.3	924	5	BX409792
8	267.2	8.9	311	2	BF855491
9	259	8.6	570	4	BI345921
10	254.4	8.5	929	5	BX409793
11	246	8.2	397	4	BG986727
12	229.6	7.7	371	1	AJ573174
13	229.4	7.6	368	2	AW425352
14	208.2	6.9	653	6	BY739930
15	197.2	6.6	465	2	BE757677
16	190.8	6.4	771	7	CK601206
17	190.6	6.4	447	2	BB848582
18	190	6.3	811	7	CK599424
19	185.8	6.2	389	5	BY088395
20	185	6.2	236	2	BF856372
21	180.6	6.0	297	1	AA771462
22	172	5.7	976	3	BC012210
23	171.6	5.7	580	4	BI343159
24	170.4	5.7	797	4	BI663166

25	167.8	5.5	708	2	AW476323
26	163.8	5.5	972	4	BI554559
27	163.6	5.5	982	4	BF976537
28	160.8	5.4	663	9	AG112651
29	159.2	5.3	237	2	BB605899
30	158.8	5.3	663	2	BF019568
31	158.6	5.3	671	2	BF727181
32	158.6	5.3	869	7	CR553384
33	156.8	5.2	657	4	BM763093
34	156.6	5.2	844	7	CF617929
35	155.8	5.2	636	1	AA557328
36	155.2	5.2	526	5	BQ554403
37	155	5.2	1058	4	BG421064
38	154.4	5.1	997	5	BQ071271
39	153.4	5.1	887	7	CN646845
40	152.8	5.1	614	6	CB372576
41	152.2	5.1	1039	4	BG419640
42	152	5.1	718	7	CN789380
43	150.2	5.0	576	5	BQ187532
44	150.2	5.0	978	7	CN646611
45	149	5.0	556	7	CO503963

ALIGNMENTS

RESULT 1
BX460671
LOCUS
DEFINITION BX460671 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
907 bp mRNA linear EST 05-MAY-2004
CS0DF017YK06 5-PRIME, mRNA sequence.
ACCESSION BX460671
VERSION BX460671.2 GI:47053688
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 907)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
On May 22, 2003 this sequence version replaced gi:31023260.
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and cloned
into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 1025.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS0DF017BF03QPI&c=1025.r.
Location/Qualifiers
1. 907
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF017YK06"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN

Query Match 28.9%; Score 867; DB 5; Length 907;
Best Local Similarity 99.2%; Pred. No. 6.1e-164;

Matches 900; Conservative 2; Mismatches 2; Indels 3; Gaps 3;									
Qy	1903	CTCCTCCAGGTGGGACACAGAGCTGGCCCTCCCTGGATGCGAGGGGTGACCCAGGA	1962						
Db	1	CTCTCCAGGTGGGACACAGAGCTGGCCCTCCCTGGATGCGAGGGGTGGACCCAGGA	60						
Qy	1963	GGCTGAGGATCTGCGATCCACACCCACTTTCAGCGGCGCTTCAGGAACAGGCGACCAT	2022						
Db	61	GGCTGAGGATCTGTCGACATCCACACCCACTTTCAGCGGCGCTTCAGGAACAGGCGACCAT	120						
Qy	2023	CGCAGAGTTCTCCCTGGTGGTTCGCGGGGCTACGCGAGGCGTGGCTGATGGACCTT	2082						
Db	121	CGCAGAGTTCTCCCTGGTGGTTCGCGGGGCTACGCGAGGCGTGGCTGATGGACCTT	180						
Qy	2083	TGCCTTTGTGGAGATGCAGGGGCGATGCTGGGGCAGGGGCCCATGTGGCCAGGATAGC	2142						
Db	181	TGCCTTTGTGGAGATGCAGGGGCGATGCTGGGGCAGGGGCCCATGTGGCCAGGATAGC	240						
Qy	2143	CTGGGCTGTCTCCAGTCCGAGGAGGAGGAGCAGGAGGCGCAGGGCTGAGTCCCGATC	2202						
Db	241	CTGGGCTGTCTCCAGTCCGAGGAGGAGGAGCAGGAGGCGCAGGGCTGAGTCCCGATC	300						
Qy	2203	GGAGGACAGCAGGAGGCCAGGGCTGAGAGCCACTGCCCCAGGTCAGTCAAGGCGTGT	2262						
Db	301	GGAGGACAGCAGGAGGCCAGGGCTGAGAGCCACTGCCCCAGGTCAGTCAAGGCGTGT	360						
Qy	2263	GCCTGAGGTGGCAGGAGCTCCACAGAGTCTCCAGAGGCCACCCCATGGGAGGACAT	2322						
Db	361	GCCTGAGGTGGCAGGAGCTCCACAGAGTCTCCAGAGGCCACCCCATGGGAGGACAT	420						
Qy	2323	CGGCGAGTCTCCCTGGTGCAGATCCGGGACCTGTCAAGTGTATGCGGAGGCGCGACAC	2382						
Db	421	CGGCGAGTCTCCCTGGTGCAGATCCGGGACCTGTCAAGTGTATGCGGAGGCGCGACAC	480						
Qy	2383	AATATCCCTGAGCATTTCCGAGGTGGACCCCGCTTACTCACTCTCAGACCTGTACGA	2442						
Db	481	AATATCCCTGAGCATTTCCGAGGTGGACCCCGCTTACTCACTCTCAGACCTGTACGA	540						
Qy	2443	TATCAAGTACCTCCCATTCGAGTTTATGATCTTCAGGAAAGTCCCCAAAGTCCGCTCAGCC	2502						
Db	541	TATCAAGTACCTCCCATTCGAGTTTATGATCTTCAGGAAAGTCCCCAAAGTCCGCTCAGCC	600						
Qy	2503	AGAGCGGCTCTCCCGCATGGCTGAGGAGGAGTGGCCGAGTTCGCGAGGCCAGCTGGCC	2562						
Db	601	AGAGCGGCTCTCCCGCATGGCTGAGGAGGAGTGGCCGAGTTCGCGAGGCCAGCTGGCC	660						
Qy	2563	CTGGCAGGTGAATCTGGGCCCCACGACAGGCTT-GGAGATCAGAGAGTTCAGGATG	2621						
Db	661	CTGGCAGGTGAATCTGGGCCCCACGACAGGCTT-GGAGATCAGAGAGTTCAGGATG	720						
Qy	2622	TGGACGGCTCTGTCGACAGGCTGCGGTGGGACGAGCGCAAGTGGTCTCTCGCGCTCAC	2681						
Db	721	TGGACGGCTCTGTCGACAGGCTGCGGTGGGACGAGCGCAAGTGGTCTCTCGCGCTCAC	780						
Qy	2682	GCAGCCTCTTCCACTTCCC-TGGGAGGCACTGGCGCTGGATGAGCTTCGAGACT-GGG	2739						
Db	781	GCAGCCTCTTCCACTTCCC-TGGGAGGCACTGGCGCTGGATGAGCTTCGAGACTGGGG	840						
Qy	2740	GCTGGCTGAGAGTGAAGGCTCCGTGGAGCACAATCTCCCGATCTCTGAGGGCAGGCC	2799						
Db	841	GCTGGCTGAGAGTGAAGGCTCCGTGGAGCACAATCTCCCGATCTCTGAGGGCAGGCC	900						
Qy	2800	GGAAGGT 2806							
Db	901	GGAAGGT 907							
RESULT 2									
BG999843/c									
LOCUS									
DEFINITION									
MR3-HN0063-030101-001-h03 HN0063 Homo sapiens cDNA, mRNA sequence.									
ACCESSION									
BG999843									
VERSION									
BG999843.1 GI:14403915									

KEYWORDS		EST.
SOURCE		Homo sapiens (human)
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS		1 (bases 1 to 622) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE		Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE		20202663
PubMed		10737800
COMMENT		Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-HN0063-030101-001-h03&t3=2001-01-03&t4=1) Seq primer: puc 18 forward High quality sequence stop: 621. Location/Qualifiers 1..622 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_lib="HN0063" /note="Organ: head normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
FEATURES		source
ORIGIN		Query Match 18.0%; Score 541; DB 4; Length 622; Best Local Similarity 98.8%; Pred. No. 2.6e-98; Matches 566; Conservative 0; Mismatches 5; Indels 2; Gaps 2; Qy 2235 CACTGCCCAAGGTCAAGGCGCTGTGCTGAGTGGCAGGCGCTCCCTGAGTGGCAGGCGTCCACCA-CGAGC 2293 Db 622 CACTGCCCAAGGTCA-TCCAAAGGCGCTGTCTGAGGTTCGACAGGCGTCCACCAAGGAGC 564 Qy 2294 TCTCCAGAGCCACCCCATGGGAGGACATCGGCGAGTCTCCCTGTGTCAGATCCCGGAC 2353 Db 563 TCTCCAGAGCCACCCCATGGGAGGACATCGGCGAGTCTCCCTGTGTCAGATCCCGGAC 504 Qy 2354 CTGTCAAGTGTATGGGAGGCGCGACACAATATCCCTGGACATTTCCGAGGTGGACCC 2413 Db 503 CTGTCAAGTGTATGGGAGGCGCGACACAATATGCTGGACATTTCCGAGGTGGACCC 444 Qy 2414 GCCTACCTCAACCTCTCAGACCTGTACGATATCAGGTACCTCCCATTCGAGTTTATGATC 2473 Db 443 GCCTACCTCAACCTCTCAGACCTGTACGATATCAGGTACCTCCCATTCGAGTTTATGATC 384 Qy 2474 TTCAGGAAAGTCCCGCAAGTCCGCTCAGCCAGAGCCGCCCTCCCGCATGGCTGAGGAGAG 2533 Db 383 TTCAGGAAAGTCCCGCAAGTCCGCTCAGCCAGAGCCGCCCTCCCGCATGGCTGAGGAGAG 324 Qy 2534 CTGGCCGAGTTCGCGAGGCCCACTGTGGCCCTGTGGCCAGGTGAACCTGGGCCCCCAGCGAGC 2593

Db 323 CTGGCCGAGTTCCTCCGAGCCACGTCGGCCCTGGCCAGGTGAACCTGGGCCCCCAGCAGGC 264
Qy 2594 CTGAGATACAGAGGAGTCAGAGATGTGACGCGCTGCTGGAGAGGCTGCGGTGGGC 2653
Db 263 CTGAGATACAGAGGAGTCAGAGATGTGACGCGCTGCTGGAGAGGCTGCGGTGGGC 204
Qy 2654 AGGAAGGCAAGTGTCTCCCGCTCAGCAGAGCTCTTCCACTTCCCTGGGAGGACCTG 2713
Db 203 AGGAAGGCAAGTGTCTCCCGCTCAGCAGAGCTCTTCCACTTCCCTGGGAGGACCTG 144
Qy 2714 CCGCTGGATGAGCTTCGAGAGCTGGGGCTGGGTGGAGAGTCAAGGCTCCCGTGGAGCAC 2773
Db 143 CCGCTGGATGAGCTTCGAGAGCTGGGGCTGGGTGGAGAGTCAAGGCTCCCGTGGAGCAC 84
Qy 2774 ATCTCCCGATCTCGAAGGAGGCGCGGAAGGT 2806
Db 83 ATCTCCCGATCTCGAAGGAGGCGCGGAAGGT 51

RESULT 3

CK357545 723 bp mRNA linear EST 23-DEC-2003.
LOCUS AGNCOURT 17184642 NIH MGC 233 Rattus norvegicus cDNA clone
IMAGE:7105047 5', mRNA sequence.

ACCESSION CK357545
VERSION CK357545.1 GI:40323477

KEYWORDS

EST.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 723)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLNL4963 row: 1 column: 13

High quality sequence stop: 724.

Location/Qualifiers

1..723

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="IMAGE:7105047"

/tissue_type="heart, pooled"

/lab_host="DH10B Tona"

/clone_lib="NIH MGC 233"

/note="Organ: heart; Vector: pExpress-1; Site 1: EcoRV;

Site 2: NotI; RNA obtained from pooled heart tissue from a

mix of male and female animals at 8 wk old. Tissues were

snap-frozen and kept at -80C for two days before RNA

extraction and purification (Tri-reagent method). cDNA was

primed using oligo-dT primer:

5'-pGACTAGTCTAGTCGCGAGCGCGCC(T)25-3' and cloned into

the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb

resulted in an average insert size of 2 kb. This primary

library is not normalized (normalized primary library is

NIH MGC 234) and was constructed by Express Genomics

(Frederick, MD). Note: this is a NIH_MGC library."

ORIGIN

Query Match 16.1%; Score 482.2; DB 7; Length 723;
Best Local Similarity 86.5%; Pred. No. 1.8e-86;
Matches 532; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
Qy 189 AGGTCAAGGTCTACATCCAGCAGCTGGTGGAGGGCTGCACTACCTGCACAGCCATGGCG 248
Db 108 AGGTCAAGGTCTATATCCAGCAGCTGGTGGAGGGCTACACTACCTGCACAGCCATGATA 167
Qy 249 TTCTCCACCTGGACATAAAGCCCTTAACATCTCTGATGGTGCATCTCTCCCGGGAGACACA 308
Db 168 TCCTCCATCTGGACATAAAGCCCGCCCAACATCTCTGATGGTCCACCCAGCTCGGGAAGACA 227
Qy 309 TTAATACTCGGACTTTGGCTTTGCCAGAGACATACCCAGCAGAGCTGCAGTTTCAGCC 368
Db 228 TTAAGATCTGTGACTTTGGCTTTGCCAGAGATACCCCGTCAGAGCCACAGTACAGCA 287
Qy 369 AGTACGGCTCCCTGAGTTCTCTCCCGGAGATCATCAGCAGAACCTCTGAGCGAAG 428
Db 288 AGTATGCTACCTGAATTCTGTCTCCAGAGATCATCAGCAGAGTCTCTGTGAGTGAGG 347
Qy 429 CCTCGACATTTGGGCGCATGGGTGTCTCTACCTCAGCTGACCTGCTCATCCCAT 488
Db 348 GCTCAGACATCTGGGCCATGGGGCTCATCTCTACCTCAGCTCAGCTGTTTCATCCCAT 407
Qy 489 TTGCGGGGAGAGTACCGTGGCCACCTCTTGAAGCTCTCGAGGGCGCGTGTCTGTA 548
Db 408 TCCTGGAGAGAGTACCGTGGCCACCTCTGCTCAATGTTTGGAGGGCGGGTCTCTTGA 467
Qy 549 GCAGCCCCATGGCTGCCACCTCAGCGAAGACGCGCAAGAGCTTTCATCAAGGCTACGCTGC 608
Db 468 GCAGTCCCATGGCTGCCACCTCAGTGGAGATGCCAGGACTTTCATCAAGGCCACACTGC 527
Qy 609 AGAGAGCCCTTCAGGCGCGGCTAGTGGGCGGCTCTCTCCACCCCTGTTCTCTGTA 668
Db 528 AAAAGACCCCGAGGGCGGCTAGTGTCTTCCAGTGTCTTGTCTCAGCTGTTCTGTA 587
Qy 669 AATCCATGCTCGGAGGAGGCGGCTTTCATCAACACCAAGCAGCTCAAGTTCCTCTG 728
Db 588 AGTCCATGCTCGTGGAGGAGGCGGCTTTCATCAACACCAAGCAGCTCAAGTTCCTCTG 647
Qy 729 CCCGAAGTCGCTGGCAGCGTTTCCCTGATGAGCTACAAGTCCATCTCTGATGAGCTCCA 788
Db 648 CTCGAGTCTGCTGGCAGCGTTTCTTGTATGAGCTACAGTCTATCTCTGATGAGCTCCA 707
Qy 789 TCCCTGAGCTGCTGC 803
Db 708 TCCCTGAGCTGCTGC 722
RESULT 4
BG989614/c
LOCUS BG989614
DEFINITION MR2-HT1160-190101-013-e11 HT1160 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG989614
VERSION BG989614.1 GI:14393684
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zagó,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bais,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.P., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.

```

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t2=MR2-H71160-
190101-013-ell&t3=2001-01-19&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 355.

FEATURES
    source
        1..517
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /dev_stage="Adult"
            /clone_lib="HT1160"
            /note="Organ: head neck; Vector: puc18; Site 1: SmaI;
            Site 2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the pUC 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."

ORIGIN
Query Match      14.5%; Score 434; DB 4; Length 517;
Best Local Similarity 94.5%; Pred. No. 8.8e-77;
Matches 482; Conservative 0; Mismatches 25; Indels 3; Gaps 3;

Qy 266 AAGCCCTTAACATCTGTGATGTCATCTGCGCGGGAAGACATTAAATCTGGACTTT 325
    |||||
Db 508 AAGCCCTTAACATCTGTGATGTT-CATCTTTCCTCCGCGGACACATTAAATCTGGACTTT 450
    |||||

Qy 326 GGCTTTGCCAGAACATCACCACAGCAGAGCTGCAGTTTCAGCCAGTACGGCTCCCTTGAG 385
    |||||
Db 449 GGCTTTGCCAGAACATCACCACAGCAGAGCTGCAGTTTCAGCCATACGGCTCCCTTGAC 390
    |||||

Qy 386 TTGGTCT-CCCCGAGATCATTCAGCAGAACCCCTGTGAGCGAAGCCTCCGACATTTGGGC 444
    |||||
Db 389 TTGGTCTCCCCCGAGATCATTCAGCAGAACCACTGTGAGCGAAGCCTCCGACATTTGGGC 330
    |||||

Qy 445 CATGGGTGTATCTCTACCTCAGCCTGACTGCTCATCCCCATTTTCCCGCGAGAGTGA 504
    |||||
Db 329 CATGGGTGTATCTCTACCTCAGCCTGACTGCTCATCCCCATTTTCCCGCGAGAGTGA 270
    |||||

Qy 505 CCGTGCCACCTCTCTGAACGTCTTGAGGGGCGGTGTCTATGGAGCAGCCCATGGCTGC 564
    |||||
Db 269 CCGTGCCACCTCTCTGAGCGTCTTGAGGGGCGGTGTCTATGGAGCAGCCCATGGCTGC 210
    |||||

Qy 565 CCACCTCAGCGAAGACGCCAAGACTTCATCAAGGCTACGCTGCAGAGAGCCCTCAGGC 624
    |||||
Db 209 CCACCTCAGCGAAGACGCCAAGACTTCATCAAGGCTACGCTGCAGAGAGCCCTCAGGC 150
    |||||

Qy 625 CCGGCTAGTGCAGCCAGTGCCTCTCCACCCCTGCTTCTGAAATCCATGCTCGGGA 684
    |||||
Db 149 CCGGCTAGTGTGGCCAGTGCCTCTCCACCCCTGCTTCTGAAATCCATGCTCGGGA 90
    |||||

Qy 685 GGAGGCCCACTTCATCAACACCAAGCAGCTCAAGTTCCTCTGTGCGCCGAAAGTCCCTGGCA 744
    |||||
Db 89 GGAGGCCCACTTCATCAACACCAAGCAGCTCAAGTTCCTCTGTGCGCCGAA-TCACCTGGCA 31
    |||||

Qy 745 GCGTTCCTGATGAGCTACAGTCCATCCT 774
    |||||
Db 30 GCGTTCCTGATGAGCCCAAGGATCCT 1
    |||||

RESULT 5
BF160452
LOCUS

```

```

601768424F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3988060 5',
mRNA sequence.
ACCESSION      BF160452
VERSION        BF160452.1 GI:11040468
KEYWORDS       EST.
SOURCE          Mus musculus (house mouse)
ORGANISM       Mus musculus
REFERENCE      1 (bases 1 to 920)
AUTHORS        NIH-MGC http://mgs.nci.nih.gov/
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgabs-r@mail.nih.gov
                Tissue Procurement: Gilbert Smith, Ph.D.
                CDNA Library Preparation: Life Technologies, Inc.
                CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LINL at:
                http://image.llnl.gov
                Plate: LL9195 row: 0 column: 05
                High quality sequence stop: 723.

FEATURES
    source
        1..920
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C28H II"
            /db_xref="taxon:10090"
            /clone_lib="IMAGE:3988060"
            /tissue="tissue type="spontaneous tumor, metastatic to mammary.
            Stem cell origin."
            /lab_host="DH10B"
            /clone_lib="NCI_CGAP_Lu29"
            /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI;
            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
            Library constructed by Life Technologies. Investigator
            providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match      10.2%; Score 307; DB 2; Length 920;
Best Local Similarity 74.4%; Pred. No. 3.5e-51;
Matches 474; Conservative 0; Mismatches 140; Indels 23; Gaps 6;

Qy 2277 GGCTCCACACAGAGCTCTCCAGAGCCACCCCATGGGAGGACATCGGGCAGGTCTCCC 2336
    |||||
Db 3 GGGTTCCCTAAGGACCTCTCCAGGCTCACCCCTTGGGAGGAAGTTGACAGGTTTCCC 62
    |||||

Qy 2337 TGGTGAGATCCGGGACCTGTCTAGGTGATGCGAGGCGCGGACACAATATCCCTGGACA 2396
    |||||
Db 63 TGGTGAGATCCGGGATCTGTCTGGTGTGTCGGAAGCAGCTGACATATCTCTTGGACA 122
    |||||

Qy 2397 TTTCCAGGTGGACCCCGCTACTCAACCTCTCAGACCTGTACGATATCAAGTACCTCC 2456
    |||||
Db 123 TTTCTGAGTGCATCTCTGCGCTACTCAACCTCTCGGATCTATATGACATCAAAATATCTCC 182
    |||||

Qy 2457 CATTCAGATTTATGATCTTCAGGAAGTCCCAAGTC---CGCTCAGCCAGAGCGCCCT 2513
    |||||
Db 183 CATTTGAGTTCATGATCTTCAGGAAGGTGCCAAACCTATAGAGCAGCAGAGTCACTG 242
    |||||

Qy 2514 CC-----CCCATGGCTGAGGAGAGTGGCCGAGTTTCCCGAG--CCCACTGGCCCTG 2565
    |||||
Db 243 GCTCAGAAACCGAAGCTGGGCAAGGCTGGCAGATTTCCTGGAGGAGCGGCATGGCCCT 302
    |||||

Qy 2566 GCCAGGTGAATCTGGGCCCCCACCAGCGCTGGGAGATCACAGAGAGTCACTG 2617
    |||||
Db 303 GGCAGCGAGCTGGGACTTCGTCTGTGGAGATTACAGAGAGCGCGAGAGAGCCAGG 362
    |||||

Qy 2618 GATGTGACGCGCTGTCTGGCAGAGGCTGCGTGGGAGGAGCGCAAGTGTCTCGCCG 2677
    |||||
Db 363 GACCTGGAAGCATGTCTGGCGAGGCTGCTGTGGCAGGAAGCGCAAGTGTGTCC---CCC 419
    |||||

```



```
/lab host="DH10B"
/clone_lib="MARC 2PIC"
/note="vector: pCMV SPORT6; Site1: NotI; Site2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

ORIGIN
Query Match      8.6%; Score 259; DB 4; Length 570;
Best Local Similarity 69.5%; Pred. No. 1.5e-41;
Matches 405; Conservative 0; Mismatches 130; Indels 48; Gaps 2;

Qy 1900 GGGCTCTCCAGGTGGGACAGAGCCCTGGCCCTCCCTGGATGGGAGGCTGGACCA 1959
Db 27 GGCAGAGCCCAATAGCAGAGAGCCAGCCCTCCCTGGATGCTGAGGCTTACCAC 86
Qy 1960 GGAGGCTGAGGATCTGTCGACTCCACACCACCTTGACGGGCTCAGGAACAGGGAC 2019
Db 87 GGAGAGCGAAGACCTGTCTAGCTTTCACCTGGTCCACAGCGGCTCAGGAACAGGGAC 146
Qy 2020 CATGGCAGAGTTCTCCCTGGTGGTGGGGGCTACGACGCTGGTGGCTATGGAC 2079
Db 147 CACACGGAAGTTCTCTTGGGATACCGTGGGGGCTATGACGGGTTGGGGCTACGGAC 206
Qy 2080 CTTTGCCCTTTGCTGAGATGAGGGGCTATGCTGGGGCAGGGCCATGTCGGCCAGGAT 2139
Db 207 CTTTGCCCTTTGCTGGGATGCTGGGGCATGCTGGGGCAGGACCTGTGGGCCAAGAT 266
Qy 2140 AGCTTGGGCTGTGCCAGTCCGAGGAGGAGGAGCAGGAGGAGCCAGGCTGAGTCCCA 2199
Db 267 GGCCTGGGCCCTCTCAGTC----- 287
Qy 2200 GTCGAGGAGCAGAGAGGAGGAGGCTGAGAGCCCACTGCCCCAGGTCACTGCAAGGCC 2259
Db 288 CTCAGAGAGCAGAGAGGACCAAGGGGCCAGAGCCACACCCAGGCGACACAGTGCC 347
Qy 2260 TGTGCTGAGGTGCGAGGGTCCACACAGAGGCTCTCAGAGGCCACCCCATGGAGGA 2319
Db 348 CTTACTGAGGTTGGCAAGGGCCCTCTGAGGGGCTCCAGGANTGAGCTCATGGAGGA 407
Qy 2320 CAT-----CGGGCAGGTCTCCCTGGTGCAGATCCGGACCTGTGAGGTGTCGGA 2370
Db 408 CTTAAGTGGAGGCTCCCAAGGTGCTGCTGTACAGATCGGGACCTGTCTCATGGACCCGGA 467
Qy 2371 GCGCGCCGACACATATCCCTGGACATTTCCGAGGTGAGACCCGCTACCTCACTCTC 2430
Db 468 AGCAGCTGACACTGTATCCCTGGATATCTCAGAGGTGGATCTGCTCACTCAACCTGTC 527
Qy 2431 AGACCTGTACGATATCAAGTACCTCCCATTCGAGTTTATGATC 2473
Db 528 TGACCTGTATGACATCAANTACCTGCCATTTGAATTATGATC 570

RESULT 10
LOCUS BX409793 929 bp mRNA linear EST 01-MAY-2004
DEFINITION BX409793 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
ACCESSION CS0DF017YK06 5-PRIME, mRNA sequence.
VERSION BX409793
KEYWORDS EST.
SOURCE BX409793.2 GI:46926497
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 929)
Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 13, 2003 this sequence version replaced gi:30640364.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 1025.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS05BAP014ZC08_AF01308_2&c=1025.r

FEATURES
source
Location/Qualifiers
1..929
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF017YK06"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match      8.5%; Score 254.4; DB 5; Length 929;
Best Local Similarity 94.1%; Pred. No. 1.4e-40;
Matches 273; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

Qy 2712 TGCCGCTGGATGACGCTGCGAGGCTGGGGCTGCGTGAGAGAGTGAAGSCCTCCGTGGAGC 2771
Db 1 TGCCGCTGGCGAACCCTGCANGACTGGGGCTGCGTGNGNGTGAAGSCCTCCGTGGAGC 60
Qy 2772 ACATCTCCCGATCTTGAAGGGCAGGCGGAGAGTCTGGAGAAGAGGGGCCCCCAGGA 2831
Db 61 ACATCTCCCGATCTTGA--GGGCAGGCGGAGGCTCTGGAGAAGAGGGGCCCCCAGGA 118
Qy 2832 AGAAGCAGGCTTGGCTTCCTTCGGGCTCTCAGGCTCTGAAGAGCTGGGACCGAGGCCGA 2891
Db 119 AGAAGCAGGCTTGGCTTCCTTCGGGCTCTCAGGCTCTGAAGAGCTGGGACCGAGGCCGA 178
Qy 2892 CATTCTTAAGGGAGCTCTCAGATGAGACTGTGGTCTGGGGCAGTCACTGACACTGGGCT 2951
Db 179 CATTCTTAAGGGAGCTCTCAGATGAGACTGTGGTCTGGGGCAGTCACTGACACTGGGCT 238
Qy 2952 GCCAGGTGTGAGCCCGAGCCAGCTCCCGAGGCCACCTCGAGCAAAAGACGGA 3001
Db 239 GCCAGGTGTGAGCCCGAGCCAGCTCCCGAGGCCACCTCGAGCAAAAGACGGA 288

RESULT 11
LOCUS BG986727 397 bp mRNA linear EST 13-JUN-2001
DEFINITION BG986727 PM1-HT1170-030101-001-g12 HT1170 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG986727
VERSION BG986727.1 GI:14390797
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 397)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P.,
Goldman, G.H., Carvalho, A.F., Mateukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202653
PUBMED 10737800
```

COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PMI&t2=PM1-HT1170-030101-001-g12&t3=2001-01-03&t4=1) Seq primer: puc 18 forward High quality sequence start: 4 High quality sequence stop: 320. Location/Qualifiers 1. .397 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_lib="HT1170" /note="Organ: head neck; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196, 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
FEATURES	source
FEATURES	1. .397 Location/Qualifiers /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="HSPD46382" /sex="female" /tissue_type="pectoral muscle (after mastectomy)" /clone_lib="HM3/RH2"
ORIGIN	
Query Match	7.7%; Score 229.6; DB 1; Length 371;
Best Local Similarity	94.4%; Pred. No. 1.2e-35;
Matches	238; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY	2277 GGCTCCCAACAGGAGCTCTCCAGAGCCCAACCCATGGAGGACATCGGGCAGGTCTCCC 2336
Db	9 GAGTGGCACAGGTGAATGACAGAGGCCCAACCCATGGAGGACATCGGGCAGGTCTCCC 68
QY	2337 TGGTCAGATCCGGACCTGTGAGTGATGCGGAGCGCGGCACAAATATCCCTGGACA 2396
Db	69 TGGTCAGATCCGGAGCCTGTGAGTGATGCGGAGCGCGGCACAAATATCCCTGGACA 128
QY	2397 TTTCCAGAGTGACCCCGCTACTCAACTCTCAGACCTGTACGATATCAAGTACTCC 2456
Db	129 TTTCCAGAGTGACCCCGCTACTCAACTCTCAGACCTGTACGATATCAAGTACTCC 188
QY	2457 CATTCAGTTTATGATCTTTCAGGAAAGTCCCAAGTCGCGTACGAGCGCCCTCCC 2516
Db	189 CATTCAGTTTATGATCTTTCAGGAAAGTCCCAAGTCGCGTACGAGCGCCCTCCC 248
QY	2517 CCATGCTGAGG 2528
Db	249 CCATGCTGAGG 260
RESULT 13	
AW425352	368 bp mRNA linear EST 25-APR-2001
LOCUS	57189 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION	AW425352
ACCESSION	AW425352
VERSION	AW425352.1 GI:6953299
KEYWORDS	EST.
SOURCE	Bos taurus (cow)
ORGANISM	Bos taurus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos. 1 (bases 1 to 368) Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, W.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perle, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W. Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle Genome Res. 11 (4), 626-630 (2001) 21180013 11282978
REFERENCE	
AUTHORS	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
COMMENT	
COMMENT	
Query Match	8.2%; Score 246; DB 4; Length 397;
Best Local Similarity	92.4%; Pred. No. 6.1e-39;
Matches	281; Conservative 0; Mismatches 20; Indels 3; Gaps 2;
QY	30 ACAGGAGCGAGACATCTGCGCGCTGAGCCACCGCTGGTCAACGGGGTGTGGAAC 89
Db	22 ATACAGGAGCGAGATCTGCGCGCTGAGCCACCGCT-GTCAAGGAGCTGTGGAAC 80
QY	90 AGTTTGAGACCGCAAGACCTCATCTCATCTCTGGAGCTGTGCTATCCGAGGAGCTGC 149
Db	81 AGTTTGAGACCGCAAGACCTCATCTCATCTCTGGAGCTGTGCTATCCGAGGAGCTGC 140
QY	150 TGGACCGCTCTGACGAGGAGCGCTGTGAGCGAGGCGGAGTCAAGTCTTACATCCAGC 209
Db	141 TGGACCGCTCTGAGGAGGCGCTGTGAGCGAGGCGGAGTCAAGTCTTACATCCAGC 200
QY	210 AGCTGTTGAGGGGTGCACTACTGTCACGCCATGGCGTTCTCCACCTGGACATAAGC 269
Db	201 AGCTGTTGAGGGGTGCACTACTGTCACGCCATGGCGTTCTCCACCTGGACATAAGC 260
QY	270 CCTCTAACATCTGATGTGATCTCTGCGCGG--AAGACATTAAATCTCGACTTTGG 327
Db	261 CCTCTAACATCTGATGTGATCTCTGCGCGGGAAGACATTCAATCTCGACTTTTG 320
QY	328 CTTT 331
Db	321 GATT 324
RESULT 12	
AJ573174	371 bp mRNA linear EST 28-JUL-2003
LOCUS	AJ573174 HM3/RH2 Homo sapiens cDNA clone HSPD46382, mRNA sequence.
DEFINITION	AJ573174
ACCESSION	AJ573174
VERSION	AJ573174.1 GI:33297035
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 371) Laveder, P., De Pitta, C., Vitulo, N., Valle, G. and Lanfranchi, G.

Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithemall.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 20
 and -minmatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCAGCAGC
 Plate: 28 row: F column: 16
 Seq primer: ATTAGTGTCACACTATAG.

FEATURES

source

Location/Qualifiers
 1. 368
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 2B0V"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: Sall;
 Library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."

ORIGIN

Query Match 7.6%; Score 229.4; DB 2; Length 368;
 Best Local Similarity 81.3%; Pred. No. 1.3e-35;
 Matches 282; Conservative 0; Mismatches 56; Indels 9; Gaps 1;
 QY 2588 GCAGCGCTGGAGATCACAGAGAGTGCAGAGATGTGGACGGCTCTCTGGCAGAGCGTCC 2647
 Db 31 GCCAGCTGGAGATCACAGAGAGCCAGAGATGGAGGCGCTCTCTGGGAAAGCTGCC 90
 QY 2648 GTGGCAGGAAGCGCAAGTGTCTTCGCCGTGCAGCAGCTCTTCCACTTCCCTGGGAGG 2707
 Db 91 GCGCGCAGGAAGCGCAAGTGTCTTCGCCGTGCAGCAGCTCTTCCACTTCCCTGGGAGG 150
 QY 2708 CACTTCCGCTGGATGACCTGCAGAGTCTGGGCTGCTGTGAGAGTGAAGCCTCCGTG 2767
 Db 151 CACTCAGCTTCTGGAGGAGCCACAGAGTCTGGGCTACGCCAGAGGTTGAAGCCTCGGTA 210
 QY 2768 GAGCAGATCTCCCGATCTGAAGGCGAGCGGAGGCTCTGGAGAGGAGGCGGCCCC 2827
 Db 211 GCCACATCTCCGCTCTCTGAGGCGAGGCTCTGAGAGTCTGGAGAGGAGTCTCTCC 270
 QY 2828 AGGAAGAGCGAGCGCTTCTTCTTCCGCTCTCAGGTCTGAAGAGCTGGGACCGAGCG 2887
 Db 271 CGGAAGAGGCGAGCGCTTCTTCTTCC-----GCCTGAAGAGCAAGGACCGAGCA 321
 QY 2888 CCGACATCTCTAAGGAGCTCTCAGATGAGACTGTGCTCTGGGCGCA 2934
 Db 322 CCGTCTTCTAAGGAGCTCTCGATGAGAGCGTGTGTTCTGGGCGCA 368

RESULT 14

BY739930 653 bp mRNA linear EST 17-DEC-2002
 LOCUS BY739930 RIKEN full-length enriched, pooled tissues, 16 days
 DEFINITION embryo, etc. Mus musculus cDNA clone 1920063N21 5', mRNA sequence.

ACCESSION BY739930

VERSION BY739930.1

KEYWORDS GI:27163955

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 653)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuura, H.,
 Batalov, S., Beisel, K. W., Blake, J. A., Bratt, D., Brusic, V.,

Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J. C., Reid, J., Ring, B. Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yangisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354583
 12466851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Fukuda, S., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
 Adachi, J., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
 Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
 Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
 Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
 Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
 Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details
 Location/Qualifiers
 1. 653
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone_lib="RIKEN full-length enriched, pooled tissues, 16
 days embryo, etc."
 /note="pooled tissues: (dev stage=16 days
 embryo, tissue type=heart, sex=mix), (dev stage=16 days
 embryo, tissue type=kidney, sex=mix), (dev stage=17 days
 embryo, tissue type=heart, sex=mix), (dev stage=17 days
 embryo, tissue type=stomach, sex=mix), (dev stage=17 days

FEATURES
source

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 16:50:48 ; Search time 4029.45 Seconds
(without alignments)
10400.615 Million cell updates/sec

Title: US-10-077-130-6_COPY_13600_14700

Perfect score: 1101

Sequence: 1 cacagcagccactgtgac.....agcaggaaggcccatgttc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gssi:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145	13.2	479	CEI179805	CEI179805 tigr-gss-
2	57.6	5.2	932	CNS0072Q	AL066742 Drosophila
3	56.6	5.1	1218	CL081011	CL081011 CH216-160
4	56	5.1	1242	CD509648	CD509648 CDA97-D11
5	54.8	5.0	925	CNS0091P	AL053013 Drosophila
6	54.8	5.0	975	AG095351	AG095351 Pan trogl
7	53.8	4.9	925	CNS0091P	AL053013 Drosophila
8	53.4	4.9	354	BF353892	BF353892 IL5-HT070
9	52.6	4.8	462	BI976924	BI976924 486128 MA
10	52	4.7	910	CNS0060N	AL065629 Drosophila
11	50.2	4.6	510	BX478907	BX478907 DKF2p686N
12	50	4.5	932	CNS0072Q	AL066742 Drosophila
13	49.6	4.5	784	AG161654	AG161654 Pan trogl
14	49	4.5	500	BF073047	BF073047 219598 MA
15	48.8	4.4	776	CNS010RY	AL099352 Drosophila
16	48.8	4.4	870	CNS010RY	AL316153 Tetraodon
17	48.6	4.4	799	BO435439	BO435439 AGENCOURT
18	48.6	4.4	971	BQ222099	BQ222099 AGENCOURT
19	48.6	4.4	988	BE876681	BE876681 601488245
20	48.6	4.4	1046	BQ643604	BQ643604 AGENCOURT
21	48.6	4.4	1183	CR720606	CR720606 Tetraodon
22	48.6	4.4	1417	AY412116	AY412116 Homo sapi
23	48	4.4	676	CE563102	CE563102 tigr-gss-
24	48	4.4	1033	CG756677	CG756677 P051-4-D0

25	47.6	4.3	566	4	BG786337	BG786337 SEAUMC006
26	47.6	4.3	895	2	BE731455	BE731455 601567740
c 27	47.6	4.3	1014	9	CNS042T9	AL314694 Tetraodon
28	47.4	4.3	1116	4	BG810038	BG810038 mgct002x1
c 29	47.2	4.3	808	9	AG030609	AG030609 Pan trogl
30	47.2	4.3	964	9	CNS003WG	AL065254 Drosophila
c 31	47	4.3	1051	5	BUS40105	BUS40105 AGENCOURT
c 32	47	4.3	1101	9	CNS017SY	AL108460 Drosophila
c 33	46.8	4.3	1136	8	CC292742	CC292742 CH261-185
34	46.8	4.3	1286	9	AG448422	AG448422 Mus muscu
35	46.6	4.2	641	9	CE104406	CE104406 tigr-gss-
36	46.6	4.2	730	6	CA379703	CA379703 658844 NC
37	46.6	4.2	1021	5	BQ434910	BQ434910 AGENCOURT
38	46.6	4.2	4219	5	BQ033185	BQ033185 Homo sapi
39	46.4	4.2	359	5	BQ091517	BQ091517 UNM18H03
40	46.4	4.2	514	2	BE722090	BE722090 189984 MA
41	46.4	4.2	514	2	BE722099	BE722099 189998 MA
42	46.4	4.2	1178	8	CC227732	CC227732 CH261-156
43	46.4	4.2	1425	9	AG441757	AG441757 Mus muscu
c 44	46.2	4.2	613	6	CA222734	CA222734 SCEZF1403
45	46.2	4.2	836	4	BG310024	BG310024 HVSMEC001

ALIGNMENTS

RESULT 1
CEI179805
LOCUS CEI179805 479 bp DNA linear GSS 25-SEP-2003
DEFINITION tigr-gss-dog-17000326766391 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION CEI179805
VERSION CEI179805.1 GI:35328485
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
AUTHORS Kirchner, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
source
1. 479
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

Query Match 13.2%; Score 145; DB 9; Length 479;
Best Local Similarity 80.9%; Pred. No. 1.1e-21;
Matches 169; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 1 CACAGCAGCCACACTGTGACACTGTCTTGGGAGCTCCCATCGATGATGAGCGGTGTGT 60
DB 271 CAAAGTGGCCGCTCTGTGACGCTGTCTGCTGGTGGCCAGCAGCGATGTTGGCGGTGG 330

Qy	711	CCTGCCACAGGAGCAGAGGTGGC	737
Dd	238	GCGGGGCGGGGGGNNNGGGCGNC	212
RESULT 5			
CNS0091P/c	CNS0091P	925 bp	DNA linear GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TET3 end of BAC #		
DEFINITION	BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL053013		
VERSION	AL053013.1 GI:4934461		
KEYWORDS	GSS.		
SOURCE	Drosophila melanogaster (fruit fly)		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 925)		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
FEATURES	Location/Qualifiers		
source	1..925 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /clone="BACR19D16" /clone_lib="RPCI-98" /note="end : TET3"		
ORIGIN			
Query Match	5.0%; Score 54.8; DB 9; Length 925;		
Best Local Similarity	13.5%; Pred. No. 0.16;		
Matches	53; Conservative 177; Mismatches 164; Indels 0; Gaps 0;		
QY	515	CGCGTGCACCTTCACAGTGCGCACAGGAGGATGACCTCACTACTGTGGAGGCCCTGGCTC	574
Dd	925	CBBSGCSGSCSBGSSSMWTSNBSBCSSSSSTSSMSSTSSMSSTSSMSSTSSMSSTSS	866
QY	575	GCTTGCCCCCGAGGAGCCAGGAGGATGACCTCACTACTGTGGAGGCCCTGGCTC	634
Dd	865	SSGTSSACVKCNASSSCGCGGMABCMWCSSSSSCGSASARGVKVRASGAGKRGGS	806
QY	635	GAAGAAGTCGATGACCGTAGCCAGCCAGCTGGACTCATTTAGGAGCTGCCAGAGG	694
Dd	805	GGASHSSSSACBSSSSSSCASWSNSSSSASRSRRSGGAGSSSSSSSSSS	746
QY	695	ACGGCGCTCGCAGCGCTTCGACAGGAGGAGGAGGTGGCACCCTCTCTCTCAAG	754
Dd	745	ASAGSVVSSASSSSSSSCSSSVSCSVASSMSCSSBSSSSASASSSSSSSASCSCCC	686
QY	755	GCTACTCCAGGCCGATGAGTGGCCCGCCTGAGATGCTGACCTCTCACACACAGCT	814
Dd	685	TSWSCSTSAASARSSSSSSSSSMSASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	626

Qy	815	CTGATGATGATGATCCCGGAGGACACCCCTTCCCTGCTACCTACCTCAAGAAGGCTGGGA	874
Dd	625	SMSSGGSGSVSASSGMSVVSSSGGRSSGSGGGGVGGSSGSSGSGSGSVCSVCSS	566
Qy	875	GCCAGGACCTCACCCTGCGGCGGCAAGGTGG	908
Dd	565	SGMCRSCSSAAAAAASCVAASCGMMCKSKSG	532
RESULT 6			
LOCUS	AG095351	975 bp	DNA linear GSS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-096G08.F, genomic survey sequence.		
ACCESSION	AG095351		
VERSION	AG095351.1 GI:16647174		
KEYWORDS	GSS.		
SOURCE	Pan troglodytes (chimpanzee)		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,Y., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	BAC end sequences of Library PTB		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 975) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. Direct Submission Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan 1-7-22 Suhiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045; Japan E-mail:suehiro@gsc.riken.go.jp, URL:http://hgpc.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170 Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.		
COMMENT	PRIMERS Sequencing: -21M13		
LIBRARY			
Vector	: pKS145		
R.Site 1	: SacI		
R.Site 2	: SacI.		
FEATURES	Location/Qualifiers		
source	1..975 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" /clone="PTB-096G08.F" /sex="male" /cell_type="lymphoblast" /clone_lib="PTB Chimpanzee Male BAC Library"		
ORIGIN			
Query Match	5.0%; Score 54.8; DB 9; Length 975;		
Best Local Similarity	46.2%; Pred. No. 0.16;		
Matches	182; Conservative 0; Mismatches 212; Indels 0; Gaps 0;		
QY	349	GAGGTGTGTGAGTTCATCTGGCACAAAGGAATGGAGCCATCCAGCCGGTGGGGTTC	408
Dd	88	GGGGGGGGGAAGCCCCCGCGCAAAGGGCGCCCGCGGGGGGGGGCGCGCGC	147
QY	409	GAGGTGTTCCTCCAGGGTCGGCAACAGATGCTGTGTGATCAAGGGCTTCACGGCAGAAC	468
Dd	148	CGGGGGGGC	207
QY	469	CAGGGCGAGTACCACTGTGGCTGGCTCAGGGCTCCATCTGCTGCTGGCTGCCACCTTC	528
Dd	208	GGGGGGCGGGGGGGCCCCCGGGGGGGGGGGCCCCCCCCCGCGCGCGCGCGCGCGCGC	267
QY	529	CAGGTGGCACTAGAGCCCAAGCCTCTGTGTGATGAGGCCCTCAGGCCACAGCTTCGCCCGCGAG	588

D _b	268	C G C G C G C C C G G C C C C C C C C C G G C C G G G C G G G C C G G G C C C G C C C G C G C G G	327
Q _y	589	G A G C C C A G A G G T G A C C T G C A C C T A C T G T G G A G G C C C T G C T C G G A A C G T C G C A T G	648
D _b	328	G C G C C G C G G G C G G C G C T G C C G C G G G C G C G G G G C A G C G G C G G C G C C C G C G G G C C C	387
Q _y	649	A G C C T G A G C C C A C G C T G G A C T C C A T T A C G A G C T G C C A G A G A G A C G C C G C C T C G C A G	708
D _b	388	C C C C C C G G G C C G C A T A G T C G T A T T T C G A A G C C C C A C C C C G G C C C C G C G G G C G G	447
Q _y	709	G C C T G C C A C A G A G G C A G A G A G T G G A C C T G	742
D _b	448	G C C C G C G G C G C G G C C G G C C G C C G G G G G C C G T G	481

RESULT 7	CNS0091P	925 bp	DNA	linear	GSS 03-JUN-1999
CNS0091P	Drosophila melanogaster	genome survey	sequence	TET3 end of BAC #	
LOCUS	BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
DEFINITION					
ACCESSION	AL053013				
VERSION	AL053013.1	GI:4934461			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster (fruit fly)				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 925)				
AUTHORS	Genoscope.				

Genoscope. Direct Submission	Genoscope - Centre National de Séquençage : Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage : BP 191 90066 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
TITLE JOURNAL	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .
COMMENT	

FEATURES	source
Location/Qualifiers	1..925
	/organism="Drosophila melanogaster"
	/mol_type="genomic DNA"
	/db_xref="taxon:7227"
	/clone="BACR19D16"
	/clone_lib="RPCI-98"
	/note="Tc1"

Query Match	Score	DB	Length
Best Local Similarity	4.9%	53.8	925
Matches	17.0%	Pred. No. 0.26	
Conservative	135	Mismatches	99
Indels	1	Gaps	1

	Qy	Qy
1	CACAGCAGCACACTGTGACACTGTCTTTGGCAGCTCCCATGATGATGAGAGCGGTGGT	60
631	CSSSSCSSSSSTSSSTSSSTSSSSSSSSSSSYTTKSTKSTASGSGSWAGGGSG	690
61	CTCTGTGGCTACCCCGTGGAGGTGAAGAGGGGGCCACAGCCAGTGCCTGTGGCAC	120
691	STGTSSSSSSSSTSTSSSVSSGSKSTSTBSGSSSSGSSSSSTSSBSSTSTSSSSS	750
121	GAGCTGTGTGGCT--GGACCCGAGTGTGTGTGGATGGCTGTGCCCGGGGAGACCTACCG	179

Db	751	SSYSSSTCSCCTCCCYSYSSSTSSSSSTGWSSTGSSSSSVGSSSDSTSTCCSCCCY	810
Qy	180	CTTCCGTGTGCGAGCTGGGCCCTGTGGGTGCTGGGGAACCGGTTCCACTGCCCCAGAC	239
Db	811	MCTCCSYMBYCTSTSCGSSSSSGKGVTKCCGCGSSSTNGMBGTSACSSSSSSC	870
Qy	240	AGTGGCGCTTGCAGAGCCACCGAAGCCTGTGCTTCCCCAGCCC	282
Db	871	SSSSVSSSKSSKSNASSSSVSSSGSSVSSNSSSSASKSSSSGVS	913

RESULT_8							
BF353892							
LOCUS	BF353892	354 bp	mRNA	linear		EST 22-NOV-2000	
DEFINITION	IJL-HT7074-290600-108-e09 HT0704 Homo sapiens cDNA, mRNA sequence.						
ACCESSION	BF353892						
VERSION	BF353892.1	GI:11312966					
KEYWORDS	EST.						
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens						

ROMO sapientis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 354)

DIAS NETO, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsumura, A., Baia, G. S., Simpson, D. H.,
Brentain, A., deOliveira, P. S., Bucher, P., Jongseneel, C. V.,
O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel.: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
<http://www.ludwig.org.br/scrips/gethtml2.p?tl1=IL5&tl2=IL5-HT0704-2306000-108-e09&tl3=2000-06-29&tl4=1>

```

seq primer: pcc 18 forward
High quality sequence stop: 341.
Location/Qualifiers
source
1..354
FEATURES
```

```

1: 33;
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0704"
/note="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 136,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

[illegible]

```

QY 347 CTGAGCTGGTAGGTCACTCTGGCAACAAGGAATGAGCGCATCCAGCCGGTGGCGCGT 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 GGGCGGGAACGCCCGTGCCTGGCTGAAGGACAGGAAGCCATCCGCAAGAGCCAGAACT 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 407 TCGAGTGTGTCCTCCAGGTCGCGACACAGATGCTGTGATCAGGCTTCACGGCAGAG 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 ATGATGTGTGTCGAGGGGACAGATGGCCATGCTGTTCATCCGCGGGGCTCGCTCAAGG 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 467 ACCAGCGCGAGTACCACTGTGGCCTGG 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 ACGCGGCGAGTACACGTGTGAGGTGG 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
LOCUS BI976924 462 bp mRNA linear EST 23-OCT-2001
DEFINITION 486128 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BI976924
VERSION BI976924.1 GI:16351329
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

```

```

REFERENCE 1 (bases 1 to 462)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,W.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-McKown,C.G., Perlea,G., Holt,I., Karanycheva,S., Liang,F.,
Quackenbush,J. and Keefe,J.W.

```

```

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
PUBMED 11282978
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

```

```

Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGTATGACCAT
BACKWARD: GTTTCCAGTCAAGCAGC
Plate: 115 row: L column: 9
Seq primer: ATTTAGGTGACACTATAG.

```

```

FEATURES
source Location/Qualifiers
1..462
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2BOV"

```

```

Query Match 4.8%; Score 52.6; DB 4; Length 462;
Best Local Similarity 54.5%; Pred. No. 0.44;
Matches 128; Conservative 0; Mismatches 104; Indels 3; Gaps 1;
QY 12 CACTGTGACATGCTCTGGCGAGTCCCATGAGTATGAGCGGTGCTCTGTGGCTA 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 CAGCATCACCTGAGCTGGGTCGGCCCGACACCCAGGACGCGGGATGAACCGCAGGGCTA 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 CCGCGTGGAGGTGAAGGAGGGGGCCACAGGCCAGTGGCGGCTGTGCCACAGAG---CTGGT 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 81 COTGTGTGAGTCTGCGCAGCTGCGGCAGCTGTCAGTGGAGCCCGTCCACGCGGGCACAGT 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 129 GCCTGGACCCAGTGTGTGTGGATGGCTGGCCCCCGGGGAGAGACTACCGTTCCTGTGT 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 GAGGGGACCACTACACGGCCAAGGGGCTGCGCCCCCGGGAGAGACTACCTGTTTCGGGT 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 189 GCGAGCTGTGGGCGCTGTGGGTGCTGGGGAAACCGTTTCACTGCCCCACAGACAGTG 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 GACGGCGTGAACGACGGGGCCCTGGCCAGCCCACTGCTCCTGGCTCGGCGGTG 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
LOCUS CNS0060N 910 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL065629
VERSION AL065629.1 GI:4944698
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

```

```

REFERENCE 1 (bases 1 to 910)
AUTHORS Genoscope.
DIRECT SUBMISSION Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoler in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

```

FEATURES
source Location/Qualifiers
1..910
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14J21"
/clone_lib="RPCI-98"
/note="end : T7"

```

```

Query Match 4.7%; Score 52; DB 9; Length 910;
Best Local Similarity 18.4%; Pred. No. 0.66;
Matches 74; Conservative 157; Mismatches 171; Indels 0; Gaps 0;
QY 29 GGGCAGCTCCCATGAGTATGAGGGGGTGGTCTCTGTGGTACCGCGTGGAGGTGAAGG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 GGGGGGGKGGGKGGGGGGGGGGGGGGGGGGTDTTWTGKSTGGGGGGSSGSSSKG 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 89 AGGGGGCCACAGCCAGTGGCGGCTGTGCCACGAGTGGTGCCTGGACCCGAGTGTGG 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 GGGGSGSYGGGGGGYGGTGGSGGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 149 TGGATGACCTGGCCCCCGGGGAGACCTACCGTTCCTGTGGCAGCTGTGGCGCTGTGG 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 CSYSSSSTGSGCGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSTSS 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 209 GTGCTGGGAACCGGTTACCTGCGCCAGACAGTGGGCTTCAGAGCCACCGAAGCCTG 268
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 645 STTBTTBTTSTYSYSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 704
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 269 TGCCTCCAGCCCTCAGCCCTGAGAGCGGCGAGTGGCAGCTGCTGAAGATGCTCTC 328
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 705 SCTSTKBYSTSTBSYBTBTBTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 764
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 329 TGGACCTTGAGTGGTGGCTGAGGCTGAGTGCATCTCGCACAAAGGAATGGAGCGCA 388
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 765 SKBSTBTSTBKSTGTBTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 824
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 389 TCCAGCCCGTGGCGGGTTCGAGGTGCTCCACAGGTCGGC 430
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 825 TSTSSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBS 866

RESULT 11
BX478907
LOCUS DXFZp686N21206_r1.686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION BX478907
ACCESSION BX478907
VERSION BX478907.1 GI:31914176
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 510)
AUTHORS Koehler,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Oeanger,A.,
Fobbo,G., Han,M. and Wiemann,S.
TITLE EST (Koehler,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
MIPS Ingoleraedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Heinrich-
Heine-University, Duesseeldorf/Germany) within the cDNA sequencing
consortium of the German Genome Project. No sl sequence available.
This clone (DKFZp686N21206) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..510
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686N21206"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/notes="Vector: piriplex2; Site_1: SfiI; Site_2: SfiI;
cDNA-collection"

ORIGIN
Query Match 4.6%; Score 50.2; DB 5; Length 510;
Best Local Similarity 53.9%; Pred. No. 1.5;
Matches 103; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 303 GGTGCACTGGTGAAGATGCTCTCTGAGGCTTGAAGTGGTGGCTGAGGCTGGTGGT 362
|||||
Db 7 GGTGGCCAGGGAGGAGCTGGAGCTGGCTGTGAGCTGTACACGGCGGGAGACCCCGT 66
|||||
QY 363 CATCTGGCACAAAGGAATGGAGCGCATCCAGCCCCGGTGGCGGTTTCGAGGTGTTCCCA 422
|||||
Db 67 GCACTGGCTGAAGGACAGAGAGGCCATCCGCAAGAGCCAGAGTATGATGTTGTTGCA 126
|||||
QY 423 GGGTCGGCAACAGATGCTGGTATCAGGGCTTTCACGGCAGAGACAGGCGGAGTACCA 482
|||||

```

```

Db 127 GGGCAGGATGCCATGCTGGTTCATCCGGGGGCTCGCTCAGGAGCGGGCGGAGTACAC 186
QY 483 CTGTGGCCTGG 493
    |||||
Db 187 GTGTGAGGTGG 197
    |||||

RESULT 12
CNS0072Q
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL066742
VERSION AL066742.1 GI:4945205
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 932)
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
1..932
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14B09"
/clone_lib="RPCI-98"
/notes="end : T7"

ORIGIN
Query Match 4.5%; Score 50; DB 9; Length 932;
Best Local Similarity 28.3%; Pred. No. 1.9;
Matches 121; Conservative 96; Mismatches 211; Indels 0; Gaps 0;

QY 45 TGATGGAGCGGTGCTCTCTGTGGCTACCGCTGAGGTGAAGAGGGGGGCACAGGCCA 104
|||||
Db 496 KGGCGGGGGGVCGRGGSGSGSGSTGGCGGKGCGCKGCTSKGCTGBGSTGCGGGG 555
|||||
QY 105 GTGGCGGCTGTGCCACGAGCTGGTCCCTGGACCCGAGTGTGTGTGGATGGCTGGCCCC 164
|||||
Db 556 SKSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 615
|||||
QY 165 CGGGGAGACCTACCGCTTCGCTGGCAGCTGTGGGCCCTGTGGGTCTGGGGAACCGGT 224
|||||
Db 616 GGGCGSGSGSGCGCGGSSCGSCGCCGCGCGCGCGCGCGSGSGSGSGSGSGSG 675
|||||
QY 225 TCACCTGCCCCAGACAGTGGCGCTTGGCAGAGCCACCGAAGCCTGTGTCCCTCCAGCCCTC 284
|||||
Db 676 CCGGSSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735
|||||
QY 285 AGCCCTGAGAGCGCGGAGGTGGGAGCTGTGTGAAGATGTCTCTCTGGAGCTTTGAGGTG 344
|||||

```

```

Db 736 SCCCCSCCCSCGCGSSSGCGSGCGSCGCGSCGCGSCGCGSCGCGSCGCGSSGC 795
Qy 345 GCGTGGAGCTGTGTAGGTTCATCTGGCACAAAGTGAAGCATCCAGCCCGGTGGCG 404
Db 796 GSGSGSGCGSSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 855
Qy 405 GTTCGAGGTGTCTCCAGGTCGCGCAACAGATGTCTGTGTATCAAGGGCTTCACGCGCAGA 464
Db 856 SCGCGGGCGSGSCGCGSCGCGCGCGCGSGSGSGSGSGSGSGSGSGSGSGSGSS 915
Qy 465 AGACACAGG 472
Db 916 GCCSCGG 923

RESULT 13
AG161654
LOCUS
DEFINITION
Pan troglodytes DNA, clone: RP43-027N15.TJ, genomic survey
sequence.
ACCESSION
AG161654
VERSION
AG161654.1 GI:16691332
KEYWORDS
GSS.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library RPCI-43
Unpublished
2
(bases 1 to 784)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimbes@sk.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
Location/Qualifiers
1..784
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-027N15.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"

FEATURES
source
Query Match 4.5%; Score 49.6; DB 9; Length 784;
Best Local Similarity 45.1%; Pred. No. 2.2;
Matches 296; Conservative 0; Mismatches 354; Indels 7; Gaps 3;

Qy 69 CTACCGGTGAGGTGAAGAGGGGGCCACAGGCCAGTGGGGTGTGCCACGAGCTGGT 128
Db 34 CTGCGGAGGGGGGGGGCGCGCGGGGNGCGCGGGGGGGCCCAAAATCAACCCCCCCC 93
Qy 129 GCCTGGACCGAGTGTGTGTGGTGTGGCTGTGGCCCCCGGGGAGACCTACCGCTTCCGTGT 188
Db 94 GCCTGCCCCCGCCCGGGGAATAGGGGGGGCGNNACGGANACC-CCGGCGGTTTTC 152

```

```

Qy 189 GGCAGCTGTGGGCGCTGTGGGTGTGGGAAACCGGTTTCACTGCCCCACAGACAGTGGGCT 248
Db 153 GCGCGCGCGCGCGAGCGGGCGCGCGGGGGGCGCGGGGGGCGGGGGGGCGGGGGCGG 212
Qy 249 TGCAGAGCACACCGAAGCCTGTGGCTTCCACAGCCCTCAGACCCCTGAGAGCCCGCAGGTGGC 308
Db 213 GNCGGGGCGCGGGGGGGGGGGGGTTCGCGCGCGCCCCCGCGCGGGGGCGGGGGCGG 272
Qy 309 AGCTGGTGAATGTCTCTCTGGAGCTTGAGGTGGTGGCTGAGGCTGGTGAAGTTCATCTG 368
Db 273 ACCGCGGGGGGGGGCGCGCGCGCTGCCCGCGCGCGCGGGGGCGCGTGTGGGGGGACCC 332
Qy 369 GCACAAGGGAATGAGCGCATCCAGCCGCGTGGCGGTTTCGAGGTGGTCTCCACAGGGTCG 428
Db 333 GAGCCCCCGCGGGCGGGCGGGCGCGCGGGGGCGGGGGCGGGGGGGGGGGGGGGGG 392
Qy 429 GCAACAGATGTGTGTATCAAGGCTTTCACGGCAGAAAGACAGGGCGAGTACCACTGTGG 488
Db 393 CGGGCGGGGGCGCGGGGGCGCGCGCGCGCGGGGGCGGGGGGGGGGGGGGGGGGG 451
Qy 489 CTTGGCTCAGGGCTCCATCTGCCCTGCGGGTGGCCACCTTCCAGGTGGCACTGAGCCAGC 548
Db 452 -CGGGCGCGGGGGCGCGGGGGCGCGACCGCGCGCGCGGGCGGGCGGGCGGGGGC 510
Qy 549 CTCTGTGATGAGGCCCTCAGCCCCAGCTTGCCTCCCGCGAGCGAGCCACAGAGGGGTGACCT 608
Db 511 GCGCGCGGGGGGGGGGGGGCGCGCGCGGGGGCGAGGGAGCGCGGGGGGGGGGGGG 570
Qy 609 GCACCTACTGTGGAGGCCCTTGGTTCGGAAGCTGTCATGAGCGCTGAGCCACAGCTGGA 668
Db 571 GCGCG---GGGGGGGGCGCGCGCGACCGCGCGGGGGGGGGGGGGGGGGGGCGG 626
Qy 669 CTCCATTAGCGAGCTGCCAGAGGAGGAGCGCGCGCTCCAGCGGCTGCCACAGGAGGC 725
Db 627 CCCCCCGCGGGGGGGGGGGGGCGCGGGGGTGGGGGGCGACCGCCCGCGGGGGC 683

```

RESULT 14

BF073047

LOCUS

219598 MARC 2BOV Bos taurus

DEFINITION

BF073047

ACCESSION

BF073047.1

VERSION

GI:10866558

KEYWORDS

EST.

SOURCE

Bos taurus (cow)

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovinae; Bos.

REFERENCE

1 (bases 1 to 500)

AUTHORS

Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,

Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C.,

Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,

Chitko-McKown, C.G., Perte, G., Holt, I., Karamycheva, S., Liang, F.,

Quackenbush, J. and Keel, J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA

libraries and construction of a gene index for cattle

JOURNAL

Genome Res. 11 (4), 626-630 (2001)

MEDLINE

21180013

PUBMED

11282978

COMMENT

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCCAAGTCAGCAGC

THIRD SHIFTS
BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:15:10 ; Search time 581.044 Seconds
(without alignments)
10198.307 Million cell updates/sec

Title: US-10-077-130-6_COPY_2000_3000
Perfect score: 1001
Sequence: 1 cgctcaccatccggagggtg.....ctgaggcagggggccaatgcc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : N_Geneseq_16Dec04:.*
1: geneseqn1980s:.*
2: geneseqn1990s:.*
3: geneseqn2000s:.*
4: geneseqn2001as:.*
5: geneseqn2001bs:.*
6: geneseqn2002as:.*
7: geneseqn2002bs:.*
8: geneseqn2003as:.*
9: geneseqn2003bs:.*
10: geneseqn2003cs:.*
11: geneseqn2003ds:.*
12: geneseqn2004as:.*
13: geneseqn2004bs:.*
```

SUMMARIES

Result No.	Score		Query Match	Length			DB	ID	Description
	Score			Match	DB	ID			
C	1	1001	100.0	24120	8	ABX11642			Abx11642 Human ser
	2	983	98.2	20489	12	ADQ22881			Adq22881 Human sof
	3	830.8	83.0	2768	6	ABX71198			Abx71198 Novel hum
	4	224	22.4	1251	10	ADC30509			Adc30509 Human nov
	5	224	22.4	2488	4	AH18453			Aah18453 Human cdn
	6	221.6	22.1	3956	13	ADR08183			Adr08183 Full leng
	7	98.4	9.8	724	4	AAH08421			Aah08421 Human cdn
C	8	98.4	9.8	2184	4	AAH18256			Aah18256 Human cdn
	9	93	9.3	1458	12	ADL24314			Adl24314 AW755252-
	10	92.4	9.2	5382	10	ADD14722			Add14722 Human src
	11	90.8	9.1	2737	4	AAI59251			Aai59251 Human pol
	12	90.8	9.1	5378	13	ADR67197			Adr67197 Human bla
	13	89.4	8.9	3999	4	AAI61037			Aai61037 Human pol
	14	88.4	8.8	1645	5	AAF24162			Aaf24162 Human sec
C	15	88.4	8.8	1948	2	ABV74346			Abv74346 Human IL-
	16	88.4	8.8	1949	2	ABV74347			Abv74347 Human IL-
	17	88.4	8.8	2170	6	ABQ54970			Abq54970 Human ova
	18	86.2	8.6	2534	11	ADM03261			Adm03261 Human cdn
	19	86.2	8.6	7564	8	ABZ24581			Abz24581 Human cel
	20	78.6	7.9	14061	6	ABV99363			Abv99363 Human nov
	20	78.6	7.9	14061	6	ABV99363			Abv99363 Human nov

ALIGNMENTS

RESULT 1

RESULTS 1
ABX11642
ID ABX11642 standard: cDNA: 24120 BP.

XX
AC ABX11642;
XX
DT 09-MAY-2003 (first entry)
XX
DE Human serine/threonine or

Human; ss; Gene; serine/threonine kinase; protein kinase; 12599;
 KW cardiovascular disease; heart failure; myocardial infarction;
 KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
 KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
 KW haemolytic anaemia; cellular proliferative disorder; cancer;
 KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
 KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
 KW multiple sclerosis

XX	Key	Location/Qualifiers
HH	5'UTR	1..71
FT		/tag= a
FT		72..23978
FT	CDS	/tag= b
FT		/product= "Kinase 1
FT		/note= "This CDS is
FT		23979..24120
FT	3'UTR	/tag= c
FT		
FT		

PT sarcoma.
 XX Example 2; SEQ ID NO 5701; 210pp; English.
 XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytosatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 SQ Sequence 20489 BP; 3917 A; 5815 C; 6983 G; 3225 T; 0 U; 549 Other;

Query Match 98.2%; Score 983; DB 12; Length 20489;
 Best Local Similarity 98.2%; Pred. No. 2.4e-191;
 Matches 983; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 QY 1 CGCTCACCATCCGGAGGTGCCCGCCAGCTGCGAGGGCGCAGTGAAGTTCGTGGCCA 60
 DB 2071 CGCTCACCATCCGGAGGTGCCCGCCAGCTGCGAGGGCGCAGTGAAGTTCGTGGCCA 2130
 QY 61 ACGCCATTGAGAGCAGCATCCGGATGAGAGTCCGGGGCGGCCCGGCTGACTGCCAACA 120
 DB 2131 ACGCCATTGAGAGCAGCATCCGGATGAGAGTCCGGGGCGGCCCGGCTGACTGCCAACA 2190
 QY 121 AGCCGCCAGCCGAGCTGCCCGGAGGTGCTGGCTGGCTGCGACGAGAGCGCAGCTGC 180
 DB 2191 AGCCGCCAGCCGAGCTGCCCGGAGGTGCTGGCTGGCTGCGACGAGAGCGCAGCTGC 2250
 QY 181 TGGCTGAGCTGCAGATCAGCTCGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTG 240
 DB 2251 TGGCTGAGCTGCAGATCAGCTCGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTG 2310
 QY 241 CCCGAGGCCCAAGTATGAGTGCAGGATCGGCGCGGGCGGGTGTCTCTTGTGGAG 300
 DB 2311 CCCGAGGCCCAAGTATGAGTGCAGGATCGGCGCGGGCGGGTGTCTCTTGTGGAG 2370
 QY 301 ATGTGGCCCGGAGCAGATCAGGCTCTACGAGTGGCTCAGCGCGGGGGCGGCATCGCCT 360
 DB 2371 ATGTGGCCCGGAGCAGATCAGGCTCTACGAGTGGCTCAGCGCGGGGGCGGCATCGCCT 2430
 QY 361 ACCAGCTCTCCGTGCAGAGGCTCGCGGCTTCTGCAAGGACATGCGGGCGAGCTGTG 420
 DB 2431 ACCAGCTCTCCGTGCAGAGGCTCGCGGCTTCTGCAAGGACATGCGGGCGAGCTGTG 2490
 QY 421 TGGATGCGCTGGCTGGGGGGCGCGCAGTTGAGTGTGAGACCTCCGAAGCCCAACGTC 480
 DB 2491 TGGATGCGCTGGCTGGGGGGCGCGCAGTTGAGTGTGAGACCTCCGAAGCCCAACGTC 2550
 QY 481 ACGTGCACTGTACAAGATGGATGAGCTGGGCGCACTCCGGTGGAGCGCTTCTTGCAGG 540
 DB 2551 ACGTGCACTGTACAAGATGGATGAGCTGGGCGCACTCCGGTGGAGCGCTTCTTGCAGG 2610
 QY 541 AGGATGTGGGACCGGCGCAGCTGTGGGAGCCACAGTACCAGGAGATGAGGCA 600
 DB 2611 AGGATGTGGGACCGGCGCAGCTGTGGGAGCCACAGTACCAGGAGATGAGGCA 2670
 QY 601 CCTACTCTCGCGGTGGCGGAGGACTCTGTGGGACTTCCGGCTCCGGCTCTGAGGCCA 660
 DB 2671 CCTACTCTCGCGGTGGCGGAGGACTCTGTGGGACTTCCGGCTCCGGCTCTGAGGCCA 2730
 QY 661 AGGTGGTGTGTTAAGGAGCAGCTGGCAGCGAGGAAGCTCAGGCGAGGAGGAGGCA 720
 DB 2731 AGGTGGTGTGTTAAGGAGCAGCTGGCAGCGAGGAAGCTCAGGCGAGGAGGAGGCA 2790
 QY 721 GTGCCACACTAGCTGCGAGGTGGCGGCCAGGCCAGACGAGGTGACGTGGTACAGGATG 780

DB 2791 GTGCCACACTAGCTGCGAGGTGGCGGCCAGGCCAGAGCGAGGTGACGTGGTACAGGATG 2850
 QY 781 GGAAGAAGCTGAGCTCCAGCTCGAAGTGTGCATGGAGGCCACAGGCTGCACGCGCAGGC 840
 DB 2851 GGAAGAAGCTGAGCTCCAGCTCGAAGTGTGCATGGAGGCCACAGGCTGCACGCGCAGGC 2910
 QY 841 TGGTTGTGCAGCAGGCGAGCCAGCGCATCCGGGGAGTATAGCTGCGAGGCTGGGGGCC 900
 DB 2911 TGGTTGTGCAGCAGGCGAGCCAGCGCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2970
 QY 901 AGCGGCTCTCTCTTCATCTCGATGTCAAAGAGCCCAAGGTGGTGTGTCACCAAGCAGG 960
 DB 2971 AGCGGCTCTCTCTTCATCTCGATGTCAAAGAGCCCAAGGTGGTGTGTCACCAAGCAGG 3030
 QY 961 TGGCACACAGTGTGAGGTGCAGGCTGAGGCGAGGGGCAATGCC 1001
 DB 3031 TGGCACACAGTGTGAGGTGCAGGCTGAGGCGAGGGGCAATGCC 3071
 RESULT 3
 ABX71198
 ID ABX71198 standard; cDNA; 2768 BP.
 XX AC ABX71198;
 XX DT 05-MAR-2003 (first entry)
 XX DE Novel human cDNA sequence #423.
 KW Human; gene; ss; nervous system disorder; peripheral neuropathy;
 KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
 KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
 KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
 KW insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
 KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
 KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
 KW coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
 KW Crohn's disease; anaphylaxis; proliferation; chemotactic;
 KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
 KW haemostatic; antiinflammatory; expressed sequence tag; EST.
 XX Homo sapiens.
 XX WO200281731-A2.
 XX 17-OCT-2002.
 XX 29-JAN-2002; 2002WO-US001222.
 XX 30-JAN-2001; 2001US-00774528.
 XX (HYSE-) HYSEQ INC.
 XX (GOOD/) GOODRICH R. W.
 XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 XX Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
 XX WPI; 2003-058563/05.
 XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid
 XX or lymphoid cell disorders, bone disorders, mechanical and traumatic
 XX disorders, coagulation disorders, and inflammatory diseases.
 XX Claim 1; Page; 612pp; English.
 CC This invention relates to the cDNA sequences encoding an isolated novel
 CC human polypeptide. The protein encoded by the nucleic acid of the
 CC invention is useful for treating central and peripheral nervous system
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)

CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)
CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
CC bacterial, viral or fungal infections; allergic conditions such as
CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
CC disease, anaphylaxis). The protein may be used to inhibit the growth,
CC infection or function of infectious agents such as bacteria, fungi,
CC viruses, or to effect bodily characteristics, biorhythms or circadian
CC cycles of rhythms. The protein may also have
CC proliferation/differentiation, stem cell growth factor, haematopoiesis
CC regulation, immune stimulation or suppressing, chemotactic/chemokinetic,
CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
CC activities. The cDNA sequences of the invention are useful for expressing
CC recombinant protein for analysis. The present sequence represents a novel
CC human cDNA sequence of the invention, this sequence is an expressed
CC sequence tag (EST) and was identified using subtractive hybridisation
XX
SQ Sequence 2768 BP; 456 A; 849 C; 995 G; 468 T; 0 U; 0 Other;

Query Match 83.0%; Score 830.8; DB 8; Length 2768;
Best Local Similarity 95.3%; Pred. No. 2.4e-160;
Matches 856; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 CGCTCACCATCCGGAGGTGCCCGCCAGCTGCACGGGGCGCAGCTGAAGTTCTGTGGCCA 60
Db 1861 CGCTCACCATCCGGAGGTGCCCGCCAGCTGCACGGGGCGCAGCTGAAGTTCTGTGGCCA 1920

QY 61 ACGGCAATTGAGACGACATCCGGATGGAGGTCCGGGGCGGCCCGAGGTGATGCCAACA 120
Db 1921 ACGGCAATTGAGACGACATCCGGATGGAGGTCCGGGGCGGCCCGAGGTGATGCCAACA 1980

QY 121 AGCCGCCAGCCGAGCTCCCGGGAGGTGCTGGCTCGGCTGCACGAGGAGCGCAGCTGC 180
Db 1981 AGCCGCCAGCCGAGCTCCCGGGAGGTGCTGGCTCGGCTGCACGAGGAGCGCAGCTGC 2040

QY 181 TGGCTGAGCTGTACATCAGGCTCGGCTGTGAGTGGCTGAAGGATGGTGCACACTGT 240
Db 2041 TGGCTGAGCTGTACATCAGGCTCGGCTGTGAGTGGCTGAAGGATGGTGCACACTGT 2100

QY 241 CCCAGGCCCCAAGTATGAGTGCAGGATCGGCGGGCGGGCGGGTCTCTTTGGCGAG 300
Db 2101 CCCAGGCCCCAAGTATGAGTGCAGGATCGGCGGGCGGGCGGGTCTCTTTGGCGAG 2160

QY 301 ATGTGGCCCGGGACGATGACGCTCTACGAGTGCCTCAGCGCGGGCGGCGCATCGCCT 360
Db 2161 ATGTGGCCCGGGACGATGACGCTCTACGAGTGCCTCAGCGCGGGCGGCGCATCGCCT 2220

QY 361 ACCAGCTCTCCGTGCAAGGCTTCGCGCTTTCTGCAACAGGACATGGCGGCGAGCTGTG 420
Db 2221 ACCAGCTCTCCGTGCAAGGCTTCGCGCTTTCTGCAACAGGACATGGCGGCGAGCTGTG 2280

QY 421 TGGATGCGGTGGTGGGGCGGGCGGAGTGTGAGTGTGAGACCTCCGAAGCCCACTCC 480
Db 2281 TGGATGCGGTGGTGGGGCGGGCGGAGTGTGAGTGTGAGACCTCCGAAGCCCACTCC 2340

QY 481 ACGTGCACTGTGTAAGGATGGCATGTGAGCTGGGCCACTCCGGTGGCGCTTCTTGCGAG 540
Db 2341 ACGTGCACTGTGTAAGGATGGCATGTGAGCTGGGCCACTCCGGTGGCGCTTCTTGCGAG 2400

QY 541 AGGATGTGGGAACCGGCAACCGGTGTGGAGCCACAGTCAACAGGACGAGATGAAGGCA 600
Db 2401 AGGATGTGGGAACCGGCAACCGGTGTGGAGCCACAGTCAACAGGACGAGATGAAGGCA 2460

QY 601 CCTACTCTCCGCGGTGGGAGGAGACTCTGTGAGTCTCCGGCTCCGGCTCTTGAGGCCA 660
Db 2461 CCTACTCTCCGCGGTGGGAGGAGACTCTGTGAGTCTCCGGCTCCGGCTCTTGAGGCCA 2520

QY 661 AGGTGTGTTTGTAAAGGAGCAGCTGGCAGCGAGGAAGCTGCAGGCGAGGAGGAGGCCA 720
Db 2521 AGGTGTGTTTGTAAAGGAGCAGCTGGCAGCGGCGTGCAGGAGGTGAGGTGGGGGCCA 2580

QY 721 GTGCCACATCAGCTGCAGGTGGCCAGCCAGACGAGGTGACGTGGGTACAGGATG 780
Db 2581 GCGCCACGCTGAGCTGTGAGGTGGCCAGGACAGATGGAGGTGACGTGGGTACAGGACG 2640

QY 781 GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGATGGAGGCCACAGGTGCACGCGCAGGC 840
Db 2641 GGAAGAAGTTCAGCTCCAGCTCGAAAGTGCACGCTGGAGGCTGTGGGCTGTATGGGAGGC 2700

QY 841 TGGTTGTGACGAGCGGACGCGGATCCCGGGGAGTATAGCTGCGAGGCTGGGG 898
Db 2701 TGGTTGTGACGAGCGGACGCGGATCCCGGGGAGTATAGCTGCGAGGCTGGGG 2758

RESULT 4
ADC30509
ID ADC30509 standard; cDNA; 1251 BP.
XX AC ADC30509;
XX DT 18-DEC-2003 (first entry)
XX DE Human novel cDNA sequence, SEQ ID NO:591.
XX KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianemic; anticoagulant; thrombolytic; vulnery;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 17; gene; ss.
XX OS Homo sapiens.
XX PN WO2003029271-A2.
XX PD 10-APR-2003.
XX PF 24-SEP-2002; 2002WO-US030474.
XX PR 24-SEP-2001; 2001US-0324631P.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX WPI; 2003-371981/35.
XX DR P-PSDB; ADC31480.
XX PT New polynucleotide and polypeptide useful for diagnosing, preventing or
XX treating conditions such as neurodegenerative diseases, anemias, platelet
XX disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
XX cancer.
XX PS Claim 1; SEQ ID NO 591; 1185pp; English.
XX SS The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention

(ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC33394). The nucleic acids and polypeptides of the invention are useful in diagnostics, drug screening, forensics, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a specifically claimed human cDNA sequence of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1251 BP; 329 A; 293 C; 363 G; 266 T; 0 U; 0 Other;

Query Match 22.4%; Score 224; DB 10; Length 1251;
 Best Local Similarity 83.6%; Pred. No. 2.2e-36;
 Matches 254; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 650 CTCTGAGCCCAAGGTGGTGTCTTAAGGAGCAGCTGGCAGCGAAGCTGCAGGCAGA 709
 Db 891 CCCAGAGCCCAAGGTGGTGTCTTAAGGAGCAGCTGGCAGCGAAGCTGCAGGCAGA 950

Qy 710 GGCAGGAGCCAGTGCACACTGAGCTGGAGTGGCCAGGCCCCAGAGCGAGGTGACGTG 769
 Db 951 GCGGGGGCCAGTGCACACTGAGCTGGAGTGGCCAGGCCCCAGAGCGAGGTGACGTG 1010

Qy 770 GTACAAGATGGGAAGAAGCTGAGCTCAGCTCGAAGTGTGCATGGAAGGCCACAGCTG 829
 Db 1011 GTACAAGATGGGAAGAAGCTGAGTTCAGCTCGAAGTGTGCATGGAAGGCCACAGCTG 1070

Qy 830 CACGCGCAGGCTGGTGTGACGAGCGAGCGAGCGGATCCCGGGAGTATAGTTCGA 889
 Db 1071 CACAGGAGGCTGGTGTGACGAGCGAGCGGATCCCGGGAGTATAGTTCGA 1130

Qy 890 GCGTGGGGCCAGCGGCTCTCTTCATCTGGATGTCAAAGAGCCCAAGGTGGTGTTCG 949
 Db 1131 GCGAGGGGCTCAGCAGCTCTCTTCGCTGCAGTGGCAGTCACTGTTGGGATGC 1190

Qy 950 CAAG 953
 Db 1191 TGAG 1194

RESULT 5
 AAH18453
 XX AAH18453 standard; cDNA; 2488 BP.
 XX AC AAH18453;
 XX XX
 DT 26-JUN-2001 (first entry)
 XX XX
 DE Human cDNA sequence SEQ ID NO:18549.
 XX XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000JP-00116126.
 XX 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.

09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 XX Claim 8; SEQ ID NO 18549; 2537pp + Sequence Listing; English.
 XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX SQ Sequence 2488 BP; 561 A; 637 C; 716 G; 574 T; 0 U; 0 Other;

Query Match 22.4%; Score 224; DB 4; Length 2488;
 Best Local Similarity 83.6%; Pred. No. 2.4e-36;
 Matches 254; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 650 CTCTGAGCCCAAGGTGGTGTCTTAAGGAGCAGCTGGCAGCGAAGCTGCAGGCAGA 709
 Db 891 CCCAGAGCCCAAGGTGGTGTCTTAAGGAGCAGCTGGCAGCGAAGCTGCAGGCAGA 950

Qy 710 GGCAGGAGCCAGTGCACACTGAGCTGGAGTGGCCAGGCCCCAGAGCGAGGTGACGTG 769
 Db 951 GCGGGGGCCAGTGCACACTGAGCTGGAGTGGCCAGGCCCCAGAGCGAGGTGACGTG 1010

Qy 770 GTACAAGATGGGAAGAAGCTGAGCTCAGCTCGAAGTGTGCATGGAAGGCCACAGCTG 829
 Db 1011 GTACAAGATGGGAAGAAGCTGAGTTCAGCTCGAAGTGTGCATGGAAGGCCACAGCTG 1070

Qy 830 CACGCGCAGGCTGGTGTGACGAGCGAGCGGATCCCGGGAGTATAGTTCGA 889
 Db 1071 CACAGGAGGCTGGTGTGACGAGCGGCGGATCCCGGGAGTATAGTTCGA 1130

Qy 890 GCGTGGGGCCAGCGGCTCTCTTCATCTGGATGTCAAAGAGCCCAAGGTGGTGTTCG 949
 Db 1131 GCGAGGGGCTCAGCAGCTCTCTTCGCTGCAGTGGCAGTCACTGTTGGGATGC 1190

Qy 950 CAAG 953
 Db 1191 TGAG 1194

RESULT 6
 ADR08183/c
 ID ADR08183 standard; cDNA; 3956 BP.

XX ADR08183;
 XX 04-NOV-2004 (first entry)
 XX Full length human cDNA useful for treating neurological disease Seq 1689.
 XX gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
 KW osteoporosis; neurological disease; Alzheimer's disease;
 KW Parkinson's disease; dementia; short memory; cancer;
 KW sense or motor function; emotional reaction; fear response; panic;
 KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;
 KW tranquilliser.
 XX Homo sapiens.
 XX EP1447413-A2.
 XX 18-AUG-2004.
 XX 12-FEB-2004; 2004EP-00003145.
 XX 14-FEB-2003; 2003JP-00102207.
 XX 09-MAY-2003; 2003JP-00131452.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
 PI Wakamatsu A, Ishii S, Nagai K, Irie R;
 XX WPI: 2004-583265/57.
 XX P-PSDB; ADR10139.
 XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
 XX Claim 1; SEQ ID NO 1689; 2686pp; English.
 XX This invention relates to novel, isolated full length human cDNA
 CC molecules and the encoded proteins thereof. Specifically, it refers to
 CC cDNA clones obtained by an oligo-capping method, where none of these
 CC clones are identical to any known human mRNAs. The present invention
 CC describes an immunoassay to identify agonists and antagonists, as well as
 CC antibodies, antisense molecules and siRNAs that can all be used to bind
 CC to and modulate expression of the cDNA molecules. As such, these
 CC molecules are useful for diagnostic markers or therapeutic targets for
 CC the various diseases or morbid states. In particular, they are useful in
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
 CC disease, Parkinson's disease, dementia, short memory and various cancers,
 CC as well as for maintaining equilibrium of sense or motor function, and
 CC for treating emotional reaction, fear response and panic. Accordingly,
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
 CC cyostatic and tranquilliser activities. This polynucleotide is a full
 CC length human cDNA sequence of the invention. NOTE: This sequence is not
 CC given in the sequence listing of the specification but can be obtained on
 CC CD-ROM from the European Patent Office, Vienna Sub-office.
 XX Sequence 3956 BP; 882 A; 1211 C; 933 G; 930 T; 0 U; 0 Other;
 Query Match 22.1%; Score 221.6; DB 13; Length 3956;
 Best Local Similarity 86.3%; Pred. No. 7.8e-36;
 Matches 245; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
 QY 647 CGTCTGAGCCCAAGGTGGTGTTCAGGACGCTGGCAGCGAGGAGTGCAGGC 706
 DB 1135 CATCCAGAGCCCAAGGTGGTGTTCAGGAGCAGCGCGAGGTGCAGGC 1076
 QY 707 AGAGCGAGGAGCCAGTGCACACTGAGCTGGAGGTGGCCAGGCCAGAGGAGTGAC 766
 DB 1075 GGAGCGGGGCGCACTGCCACGCTGAGCTGGAGGTGGCCAGGCCAGATGGAGGTGAC 1016
 QY 767 GTGTGTAACAAGATGGGAGAGCTGAGCTCCAGCTCGAAAGTGTGATGAGGCCACAGG 826
 |||||

Db 1015 ATGCTACAAGGACCGGAAGAGCTGAGCTCAAAAGTGCATGAGGCGCAGCG 956
 QY 827 CTGCACGCGCAGGCTGTTGTGACGAGGAGGCGGCGGATGCCGGGAGTATAGCTG 886
 Db 955 CTACACACGAGGCTGGTGTGACGAGGCGGCGGCGGATGCTGGGGAGTACAGCTG 896
 QY 887 CGAGGCTGGGGCGGCGGCGGCTCTCTTCCATCTCGATGTCAAAG 930
 Db 895 TGAGGCGGGGCGGCGGCGGCTGCTTCCGCTGCACGTGGCAG 852
 RESULT 7
 AAH08421
 ID AAH08421 standard; cDNA; 724 BP.
 XX
 AC AAH08421;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:5256.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 OS
 XX EP1074617-A2.
 FN
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 XX
 PS Claim 1; SEQ ID NO 5256; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesising 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention


```
Db      488 TCCAGTGCAGTT 499
|||||
RESULT 9
ADL24314
ID  ADL24314 standard; DNA; 1458 BP.
XX
AC  ADL24314;
XX
DT  03-JUN-2004 (first entry)
XX
DE  AW755252-interacting protein mPN34854 coding sequence SEQ ID NO: 13.
XX
KW  AW755252; cardiant; gene therapy; ischaemic heart disease;
KW  myocardial infarction; cardiac failure; dilated cardiomyopathy;
XX  angina pectoris; hypertrophia cordis; mPN34854; ds; gene.
OS  Unidentified.
XX
PN  WO2004019880-A2.
XX
PD  11-MAR-2004.
XX
PF  26-AUG-2003; 2003WO-US026997.
XX
PR  28-AUG-2002; 2002US-0406613P.
PR  18-SEP-2002; 2002US-0411657P.
XX
PA  (TAKE ) TAKEDA CHEM IND LTD.
XX
PI  Hensel C, Sakamoto T;
XX
DR  WPI; 2004-239111/22.
XX
PT  New protein complex, useful for preparing a composition for treating
PT  diseases e.g., ischemic heart disease, myocardial infarction, cardiac
XX  failure, dilated cardiomyopathy, hypertrophia cordis or angina pectoris.
PS  Claim 129; Fig 7; 151pp; English.
XX
CC  The present invention provides a new isolated protein complex comprising
CC  a first protein comprising AW755252 or its homologue or derivative or
CC  fragment interacting with a second protein consisting of MFHL2, mPN34854,
CC  mPRPH1, mCTEX-1, ACTN2, mACTN4 and mMRU, or a homologue or derivative or
CC  fragment. The protein complex is useful for preparing a composition for
CC  treating diseases e.g., ischaemic heart disease, myocardial infarction,
CC  cardiac failure, dilated cardiomyopathy, hypertrophia cordis or angina
CC  pectoris. The present sequence is the coding sequence of the mPN34854
CC  protein, which interacts with the AW755252 protein.
XX
SQ  Sequence 1458 BP; 294 A; 390 C; 470 G; 304 T; 0 U; 0 Other;
Query Match          9.3%; Score 93; DB 12; Length 1458;
Best Local Similarity 45.6%; Pred. No. 1.3e-09;
Matches 367; Conservative 0; Mismatches 435; Indels 3; Gaps 1;

Qy      143 GGAGGTGTCGCTCGCTGACGAGGAGGCGCAGCTGCTGCTGAGCTGTCAGATCAGGC 202
Db      546 GCATGTGGCAGCCATACGCTTAGAGCGTGTGTCTGACCTGTGAGCTGTGCCGACCTG 605
Qy      203 TGGCGCTGTGACGTGGCTGAAGGATGTCGCACACTGTCCCCAGGCCCAAGTATGAGGT 262
Db      606 GGCTGAGGTGCGCTGGACCAAGATGGGAGGAGGATGTCGAGAGCCCACTGCTCTCT 665
Qy      263 GCAGGCATCGCGCGCGGGGTGCTCTGTGTCGAGATGTGCCCGGACGATGTCAGG 322
Db      666 GGAGAAGAGACACCATCCGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 725
Qy      323 CCTCTACAGTGTCTAGCGCGCGCGCGCGCGCGCATCGCCTACAGCTCTCGTGAAGCCT 382
Db      726 CGAGTACCTGTGTGAAATCCATGATGATGTCGGCTTCTTCCATCACCCTCAGAGCC 785
```

```
Qy      383 CGCGCGCTTTCTGCACAAGGACATGGCGGGCAGCTGTGTGGATCCGCTGGCTGGGGGCC 442
Db      786 CCCTGTGCGGATCATATACCCCAAGGACAGGTGACCTTACACGCGCTGAGTTGGGATG 845
Qy      443 GGCGCAGTTTGTAGTGTGAGACCTCCGAAGCCACGTCACGTCGACTGGTACAAAGGATGG 502
Db      846 TGTGTGCTCACCTGTGAGCTGTCTAGAGAGGATGCTCTCTGTACGCTGGTACAAAGGATGG 905
Qy      503 CATGGAGCTGGGCCACTCCGCTGAGCGCTTCTTTCAGGAGGATGTCGGGACGCGGCACCG 562
Db      906 GTTAGAGGTGAGGAGA---GTGAAGCCCTGGTCTCCAGAGCATGGGCTCTGTCGCCG 962
Qy      563 GCTGTTGGCAGCCACAGTCACAGGAGGATGAAGGCACCTACTCTCCGCGCTGGGCGGA 622
Db      963 TCTGTTGTTCCTGCTGCCAGCAGAGGACGGGGGCGAGTTTGTGTGTGATGCTGGGGA 1022
Qy      623 GGACTGTGTGAGACTTCGGGCTCCGCGTCTCTGAGCCCAAGGTGGTGTGTTGCTAAGGAGCA 682
Db      1023 TGATTACAGCCTTCTTCACTGTCTGTCAAGCTCCACAGAAAGGATTGTGCACCCAGC 1082
Qy      683 GCTGGCAGCGAGGAAGCTGCAGGACAGGAGCGGAGCGGAGTGCACACTGAGCTGCGAGGT 742
Db      1083 GGCCCGATCCCTGATTTGAGTTTGGGGTCCAGGACAGCTGGAGCTACGCTGCGAAGT 1142
Qy      743 GGCCCGAGCCCGACAGCGAGGTGACGTGGTACAAAGGATGGGAAGAGCTGAGCTCCAGCTC 802
Db      1143 GGCCCGCGGTGGGTCTCAGGTGCGCTGGTACAAAGGATGGCTAGAGGTAGAGGTGTGAGA 1202
Qy      803 GAAAGTGTGATGAGGCGCACAGGCTGCAGCGCAGGCTGGTGTGTGAGCAGGAGGAGGCCA 862
Db      1203 TGCCTGTGAGCTGGGTGCTGAGGGGCTGCGCGCACTCTCACCTGCCCCACGCCAGCC 1262
Qy      863 GCGCGATGCGCGGAGTATAGTTCGAGGCTGGGGCGCAGCGCTCTCTCTTCCATCTGGA 922
Db      1263 TGAGGATGCGCGGAGTATGATGTGAGACCCGAGATGAGGCTGTCCACCTTCAACGTCAG 1322
Qy      923 TGTCAAAGAGCCCAAGGTGGTGT 947
Db      1323 CCTGGCTGAGCTTCGGTGCAGTTT 1347

RESULT 10
ADD14722
ID  ADD14722 standard; cDNA; 5382 BP.
XX
AC  ADD14722;
XX
DT  01-JAN-2004 (first entry)
XX
DE  Human src biomarker polynucleotide SEQ ID NO:116.
XX
KW  predictor set; protein tyrosine kinase activity modulator;
KW  protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
KW  gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
XX
OS  Homo sapiens.
XX
PN  WO2003062395-A2.
PD  31-JUL-2003.
XX
PF  17-JAN-2003; 2003WO-US001981.
XX
PR  18-JAN-2002; 2002US-0350061P.
XX
PA  (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI  Huang F, Fairchild CR, Lee FY, Shaw P;
XX
DR  WPI; 2003-636735/60.
DR  P-PSDB; ADD14123.
XX
PT  New polynucleotides and polypeptides for predicting the activity of
```

compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways.

Claim 2; SEQ ID NO 116; 139pp; English.

The present invention describes a predictor set comprising a plurality of polynucleotides or polypeptides whose expression pattern is predictive of the response of cells to treatment with a compound that modulates protein tyrosine kinase activity or members of the protein tyrosine kinase pathway. Also described: (1) predicting whether a compound is capable of modulating the activity of cells, comprising obtaining a sample of cells, determining whether the cells express a plurality of markers, and correlating the expression of the markers to the compound's ability to modulate the activity of the cells; (2) a plurality of cell lines for identifying polynucleotides and polypeptides whose expression levels correlate with compound sensitivity or resistance of cells associated with a disease state; and (3) identifying polynucleotides and polypeptides that predict compound sensitivity or resistance of cells associated with a disease state, comprising subjecting the plurality of cell lines to one or more compounds, analysing the expression pattern of a microarray of polynucleotides or polypeptides, and selecting polynucleotides or polypeptides that predict the sensitivity or resistance of cells associated with a disease state by using the expression pattern of the microarray. The polynucleotides and polypeptides have cytostatic activities, and can be used in gene therapy. The polynucleotides and polypeptides are useful in predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways. These may be used in determining drug sensitivity in patients to allow the development of individualized genetic profiles which aid in treating diseases and disorders (e.g. cancer) based on patient response at a molecular level. The present sequence is used in the exemplification of the present invention.

Sequence 5382 BP; 943 A; 1699 C; 1798 G; 942 T; 0 U; 0 Other;

Query Match 9.2%; Score 92.4; DB 10; Length 5382;

Best Local Similarity 46.7%; Pred. No. 2e-09;

Matches 365; Conservative 0; Mismatches 411; Indels 6; Gaps 2;

165	Qy	165	GAGGAGCGCAGCTGCTGGCTGAGCTGTACAGTACAGCTCGCGCTGTGACGTGCTGCTGAAG	224
1703	Db	1703	GAGCGGTGTGTGCTGACTGTGTGAGCTCTCAAGGGTGGACTTCCCGCGCAACCTGGTACAAAG	1762
225	Qy	225	GATGGTCCGACACATGTCCCCAGGGCCCCAAGTATCAGGTGACAGGCATCGGCCGGCGGGCGG	284
1763	Db	1763	GATGGGCAGAAAGTGGAGGAGAGCGAGTTGCTGTGTGTGAAGATGGATGGCGCGCAACAC	1822
285	Qy	285	GTGCTCTCTGTGCGAGATGTGGCCCCGGGACGATCAGGCCCTCTACGATGGCTCAGCGCGC	344
1823	Db	1823	CGTCTGATCTCTGAGGGCCAAAGTCCAGGACAGTGGCGAGTTTGAGTGCAGGACAGAA	1882
345	Qy	345	GGGGGCGCATCGCTACCAAGCTCTCCGTGTCAAGGCCCTCGCGCGCTTCTGCACACAGGAC	404
1883	Db	1883	GGGGTCTCGCGCTTCTTCGCGCTCACTGTGCCAAGATCCTCCCGTGCACATCGTGGACCCC	1942
405	Qy	405	ATGCGGGCGAGCTGTGTGGATGCGCTGGCTGGGGGGCCCGCGCAGTTTGAGTGTGAGACC	464
1943	Db	1943	CGAGAACATGTGTTCGTGTCATGCCATAACTTCCGAGTGTGTCACTGTGCGCTGTGAGGTG	2002
465	Qy	465	TCCGAAGCCCACTGCCAGTGCACGTGTGTGTGATGATGAGGATGGCATGGAGCTGGGCGACTCCCGT	524
2003	Db	2003	GACCGAGGAGCAGCCCCCTGTGCGTTGTTGTACAAGACGGGCGAGGAGGTGGAGGAGA--GT	2059
525	Qy	525	GAGCGCTCTTGACAGGAGGATGTGGGACCGGGCAGCCGGCTGGTGGGACGCCACAGTCACC	584
2060	Db	2060	GACTTCGTGTGTCTGGAGAAATGAGGGGGCCCCATCGCGCCCTGGTGTGCTGCCCGCCACCCAG	2119
585	Qy	585	AGGCAGGATGAAGGCACCTACTCTCGCGCGTGGGCGGAGGACTCTGTGGACTTCGGGCTC	644
2120	Db	2120	CCCTCAGACGGGGCGAGTTTCAGTGGCTCGCTGGAGATGAGTGTGCTCTACTTCACTGTC	2179
645	Qy	645	CGCGTCTCTGAGCCCAAGGTTGTTTGTCTAAGAGCAGCTGGCACCGCAGGAAGCTGCAG	704

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 16:50:48 ; Search time 3663.47 Seconds
(without alignments)
10400.615 Million cell updates/sec

Title: US-10-077-130-4_COPY_22500_23500

Perfect score: 1001
Sequence: 1 ggcacgtcagtgacactggc.....ctgtctcaagtcgtggacct 1001

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gsei:
9: gb_gsei2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	666	66.5	929	5	BX409793
2	655.6	65.5	683	2	AW862431
3	570.4	57.0	924	5	BX409792
4	475	47.5	475	4	BI916955
5	453	45.3	477	4	BI916956
6	415.2	41.5	556	4	BI776197
7	405.2	40.5	581	7	CR759117
8	397.4	39.7	771	7	CK601206
9	391.6	39.1	565	7	CR758263
10	374.2	37.4	768	2	BF134040
11	370.8	37.0	811	7	CK599424
12	340.8	34.0	567	4	BI345924
13	324.2	32.4	494	7	CF177813
14	324.2	32.4	653	6	BY739930
15	295	29.5	494	6	CB713368
16	293	29.3	526	5	BO554403
17	274	27.4	397	2	AW479627
18	272.8	27.3	485	6	CB727068
19	255	25.5	328	2	AW482385
20	246	24.6	279	2	BF762805
21	225.2	22.5	427	1	AI006121
22	178.6	17.8	447	2	BB484582
23	171.2	17.1	809	7	CN439177
24	170	17.0	654	9	CR182048

C	25	159.6	15.9	207	7	CR756901	CR756901
	26	156.4	15.6	3241	3	AK035543	Mus muscu
	27	150.2	15.0	597	4	BM488048	pgm2n.pko
	28	148.4	14.8	1030	4	BM552361	AGENCOURT
C	29	142	14.2	715	5	BM985339	UI-CP-EC1
	30	138.2	13.8	621	4	BG792248	UTSW H31D
	31	135.8	13.6	389	5	BY088395	BY088395
	32	133	13.3	444	7	CR747491	CR747491
	33	133	13.3	480	5	EX280322	EX280322
	34	132	13.2	731	4	BG818748	BG818748
	35	131.8	13.2	832	7	CK602844	AGENCOURT
	36	130	13.0	187	2	AW435707	AW435707
C	37	124.8	12.5	740	7	CK774333	CK774333
	38	114.8	11.5	902	4	BM018924	BM018924
C	39	111.4	11.1	773	5	EX435097	EX435097
	40	96.8	9.7	948	7	CO648953	CO648953
	41	95	9.5	878	4	BJ729693	BJ729693
C	42	92.2	9.2	551	4	BG001365	BG001365
	43	92.2	9.2	796	4	BG421646	BG421646
C	44	92	9.2	462	7	CV334914	CV334914
	45	92	9.2	582	7	CV327309	CV327309

ALIGNMENTS

RESULT 1
BX409793
LOCUS BX409793 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS00DF017YK06 5-PRIME, mRNA sequence.
ACCESSION BX409793
VERSION BX409793.2 GI:46926497
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 929)
AUTHORS Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 13, 2003 this sequence version replaced gi.30640364.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 1025.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS0BAF014ZC08_AF01308_2&c=1025.r

FEATURES

Location/Qualifiers
1..929
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF017YK06"
/tissue="FETAL BRAIN"
/dev_stages="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN

Query Match 66.5%; Score 666; DB 5; Length 929;


```

Db      541  GGGCCCAAGATCATCCCTACCAACCCAGAGACAGACAGTGTGCGGAATACGA 600
Qy      726  GGCCCTCAAGGCGCTGCGCCACCCGACCTGCGCCAGCTGACGAGCCTACCTCAGCCC 785
Db      601  GGCCCTCAGAGGCGCTGCGCCACCCGACCTGCGCCAGCTGACGAGCCTACCTCAGNCC 660
Qy      786  CCGGACCTCTGTGCTC 801
Db      661  CC-GCACCTGGTGCTC 675

```

RESULT 3

```

BX409792  BX409792  924 bp  mRNA  linear  EST 30-APR-2004
LOCUS      CS0DF017YK06 5-PRIME, mRNA sequence.
DEFINITION BX409792
VERSION    BX409792.1 GI:30640362
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens

```

REFERENCE

```

AUTHORS    Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    Contact: Genoscope
           Genoscope - Centre National de Sequencage
           2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

```

1st strand cDNA was primed with a NotI-oligo (dr) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.

This sequence belongs to sequence cluster 1025.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?e=CS0BAP014ZC08_AF01308_l&c=1025.r

FEATURES

source

```

1. 924
   Location/Qualifiers
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="CS0DF017YK06"
   /tissue_type="FETAL BRAIN"
   /dev_stage="fetal"
   /clone_lib="Homo sapiens FETAL BRAIN"
   /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
   was primed with a NotI-oligo (dr) primer. Five prime end
   enriched, double-strand cDNA was digested with Not I and
   cloned into the Not I and EcoRV sites of the pCMVSPORT 6
   vector. Library was not normalized."

```

ORIGIN

```

Query Match      57.0%; Score 570.4; DB 5; Length 924;
Best Local Similarity 95.2%; Pred. No. 4.9e-119;
Matches 641; Conservative 0; Mismatches 24; Indels 8; Gaps 5;

Qy      1  GGCCAGTCAGTGACACTGGCTGCCAGGTGTGAGCCGACGCTGCCAGGCCACCTGG 60
Db      208 GGCCAGTCAGTGACACTGGCTGCCAGGTGTGAGCCGACGCTGCCAGGCCACCTGG 267
Qy      61  AGCAAGACGGAGCCCGCTGGAGAGCAGCAGCGCTGCTCATCTTGCACCCCTCAAG 120
Db      268 AGCAAGACGGAGCCCGCTGGAGAGCAGCAGCGCTGCTCATCTTGCACCCCTCAAG 327
Qy      121 AACTTCCAGCTTCTGACCATCTCTGCTGGTGGTGGCTGAGGACCTGGGTGTGTACACCTGC 180
Db      328 AACTTCCAGCTTCTGACCATCTCTGCTGGTGGTGGCTGAGGACCTGGGTGTGTACACCTGC 387

```

```

Qy      181  AGCGTGAGCAATCGCTGGGACAGTGACACCAACGGGGCTCTCCGGAGGACAGCGC 240
Db      388  AGCGTGAGCAATCGCTGGGACAGTGACACCAACGGGGCTCTCCGGAGGACAGCGC 447
Qy      241  CCCTCATCTTCGCGATGCGCCGATATCGGGAGGTGTACGGGATGGGGTGTCTGTGTC 300
Db      448  CCCTCATCTTCGCGATGCGCCGATATCGGGAGGTGTACGGGATGGGGTGTCTGTGTC 507
Qy      301  TGGAAAGCCCGTGGAAATCTACGGCCCTGTGACCTACATTTGTGCACTAGAGCC 360
Db      508  TGGAAAGCCCGTGGAAATCTACGGCCCTGTGACCTACATTTGTGCACTAGAGCC 567
Qy      361  GGCAGCTGGACCACTGGCTCCGACATCTTGTGCTGCTCTACCTACCTACCTACCT 420
Db      568  GGCAGCTGGACCACTGGCTCCGACATCTTGTGCTGCTCTACCTACCTACCTACCT 627
Qy      421  TCCGGGGTGGCACCCTACACCTTCCGACGGCATGTGTGACGAGGAGGAAATGGGTCCC 480
Db      628  TCCGGGGTGGCACCCTACACCTTCCGACGGCATGTGTGACGAGGAGGAAATGGGTCCC 687
Qy      481  TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGCCAGCCACCTGGCTCTGAGGAG 540
Db      688  TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGCCAGCCACCTGGCTCTGAGGAG 747
Qy      541  GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCAGCAGCACAAGACCTTCGATTCAGACA 600
Db      748  GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCAGCAGCACAAGACCTTCGATTCAGACA 803
Qy      601  CAGATCCAGAGGGCGGCTTCAGCGTGTGCGGCAATCTGGGAGAGCCAGCGGGCGG 660
Db      804  CAGATCCAGAGGGCGGCTTCAGCGTGTGCGGCAATCTGGGAGAGAG--CCACGGGCGG 859
Qy      661  CGCTGGCGGCCA 673
Db      860  CGCTGTCCGACA 872

```

RESULT 4

```

BX16955  BX16955  475 bp  mRNA  linear  EST 16-OCT-2001
LOCUS      60317768F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5242025 5'
DEFINITION mRNA sequence.

```

```

ACCESSION  BX16955
VERSION    BX16955.1 GI:16180909
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

```

AUTHORS    NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-remail.nih.gov

```

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM11610 row: g column: 18

High quality sequence stop: 475.

Location/Qualifiers

FEATURES

source

```

1. 475
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="IMAGE:5242025"
   /lab_host="DH10B"
   /clone_lib="NIH_MGC_121"
   /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;

```

Site 2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 47.5%; Score 475; DB 4; Length 475;
Best Local Similarity 100.0%; Pred. No. 2.2e-97;
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GTGACACTGGCTGCCAGGTGTGAGCCAGCCAGCTGCCAGGCGCCACCTGGAGCAAGAC 69
Db 1 GTGACACTGGCTGCCAGGTGTGAGCCAGCCAGCTGCCAGGCGCCACCTGGAGCAAGAC 60

QY 70 GGAGCCCCCTGGAGAGCAGCAGCGGTGCTCATCTCTGCCACCCCTCAAGAACTTCCAG 129
Db 61 GGAGCCCCCTGGAGAGCAGCAGCGGTGCTCATCTCTGCCACCCCTCAAGAACTTCCAG 120

QY 130 CTTCTGACCATCTGGTGTGTGGCTGAGGACCTGGGTGTGTACCTGTCAGCGTGAAGC 189
Db 121 CTTCTGACCATCTGGTGTGTGGCTGAGGACCTGGGTGTGTACCTGTCAGCGTGAAGC 180

QY 190 AATGGCTGGGACAGTGTACACACAGCGGCGTCTCCGAGGCGAGCGCCCTCATCT 249
Db 181 AATGGCTGGGACAGTGTACACACAGCGGCGTCTCCGAGGCGAGCGCCCTCATCT 240

QY 250 TCGCCATGCCCGGATATCGGGAGGTGTACCGGATGGGGTGTCTGTCTGGAAGCCC 309
Db 241 TCGCCATGCCCGGATATCGGGAGGTGTACCGGATGGGGTGTCTGTCTGGAAGCCC 300

QY 310 GTGGAACTCTAGCGCCCTGTGACCTACATTGTGAGTGTGAGCTGTGAGGCGGAGCTGG 369
Db 301 GTGGAACTCTAGCGCCCTGTGACCTACATTGTGAGTGTGAGGCGGAGCTGG 360

QY 370 ACCACACTGGCTCCGACATCTTTGACTGTGCTACCTGACGACGCAAGCTCTCCCGGGGT 429
Db 361 ACCACACTGGCTCCGACATCTTTGACTGTGCTACCTGACGACGCAAGCTCTCCCGGGGT 420

QY 430 GGCACCTACACTTCCGACGCGATGTGTACGAGGAGGAGCAAGTGGGTCCCTACA 484
Db 421 GGCACCTACACTTCCGACGCGATGTGTACGAGGAGGAGCAAGTGGGTCCCTACA 475

RESULT 5

BI916956
LOCUS 603177770F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5242121 5',
DEFINITION mRNA sequence.

ACCESSION BI916956

VERSION BI916956.1 GI:16180910

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 477)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Sequencing by: The I.M.A.G.E. Consortium (LNL)

DNA Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

plate: LLAM11610 row: k column: 18
High quality sequence stop: 477.

FEATURES

source Location/Qualifiers

1. .477

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5242121"

/lab_host="DH10B"

/clone_lib="NIH_MGC_121"

/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: EcoRV (destroyed); RNA source anonymous pool of 3

fetal brains, female age 20 weeks, female age 24 weeks,

and male age 26 weeks. Library is oligo-dT primed and

directionally cloned (EcoRV site is destroyed upon

cloning). Average insert size 1.7 kb, insert size range

0.7-3.5 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 017. Note:

this is a NIH_MGC Library."

ORIGIN

Query Match 45.3%; Score 453; DB 4; Length 477;
Best Local Similarity 99.6%; Pred. No. 2.2e-92;
Matches 475; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 10 GTGACACTGGCTGCCAGGTGTGAGCCCA-GCCAGCTGCCAGGCGCCACCTGGAGCAA-AG 67
Db 1 GTGACACTGGCTGCCAGGTGTGAGCCCATGCCAGCTGCCAGGCGCCACCTGGAGCAAAG 60

QY 68 ACGAGAGCCCCCTGGAGAGCAGCGGTGTCTCTCTGCCCACCTCAAGAACTTCC 127
Db 61 ACGAGAGCCCCCTGGAGAGCAGCGGTGTCTCTCTGCCCACCTCAAGAACTTCC 120

QY 128 AGCTTCTGACATCTCGTGTGTGGTGTGAGACCTGGGTGTGTACCTGTCAGCGTGA 187
Db 121 AGCTTCTGACATCTCGTGTGTGGTGTGAGACCTGGGTGTGTACCTGTCAGCGTGA 180

QY 188 GCAATGGCTGGGAGCAGTGACACACAGCGGCGTCTCCGGAAGGAGGAGCGCCCTCAT 247
Db 181 GCAATGGCTGGGAGCAGTGACACACAGCGGCGTCTCCGGAAGGAGGAGCGCCCTCAT 240

QY 248 CTTGCGCATGCCCGGATATCGGGAGGTGTACCGGATGGGGTGTCTGTCTGGAAGC 307
Db 241 CTTGCGCATGCCCGGATATCGGGAGGTGTACCGGATGGGGTGTCTGTCTGGAAGC 300

QY 308 CCGTGGAACTCTAGCGCCCTGTGACCTACATTGTGAGTGTGAGGCGGAGCT 367
Db 301 CCGTGGAACTCTAGCGCCCTGTGACCTACATTGTGAGTGTGAGGCGGAGCT 360

QY 368 GGACACACTGGCTCCGACATCTTTGACTGTGCTACCTGACGACGAGCTCTCCCGGG 427
Db 361 GGACACACTGGCTCCGACATCTTTGACTGTGCTACCTGACGACGAGCTCTCCCGGG 420

QY 428 GTGGCACTTACCTTCCGACGCGATGTGTACGAAAGGAGGAGTGGGTCCCTACA 484
Db 421 GTGGCACTTACCTTCCGACGCGATGTGTACGAAAGGAGGAGTGGGTCCCTACA 477

RESULT 6

BI776197

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Bos taurus (cow)

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovinae; Bos.

1 (bases 1 to 556)

REFERENCE

AUTHORS

Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,

BI776197 556 bp mRNA linear EST 25-SEP-2001
468982 MARC 2BOV Bos taurus cdna 5', mRNA sequence.

BI776197

BI776197.1 GI:15777174

EST.

Bos taurus (cow)

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovinae; Bos.

1 (bases 1 to 556)

REFERENCE

AUTHORS

Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,

TITLE		Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle	
JOURNAL	Genome Res. 11 (4), 626-630 (2001)		
MEDLINE	21180013		
PUBMED	11282978		
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options. PCR Primers FORWARD: AGGAACACGCTATGACCAT BACKWARD: GTTITCCCCAGTCACGACG Plate: 93 row: C column: 9 Seq primer: ATTAGGTGACACTATAG. Location/Qualifiers 1..556 /organism="Bos taurus" /mol_type="mRNA" /db_xref="taxon:9913" /tissue_type="pooled" /lab_host="DH10B" /clone_lib="WARC 2BOV" /notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from testis, thymus, semitendosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."		
FEATURES		source	
ORIGIN			
Query Match	41.5%	Score 415.2;	DB 4; Length 556;
Best Local Similarity	85.4%;	Pred. No. 9.1e-84;	
Matches 475;	Conservative 0;	Mismatches 78;	Indels 3; Gaps 1;
QY	118	AAGAACTTCCTCAGCTTCTGACCATCCTCGTGGTGGTGGTGTGAGGACCTGGGGTGTGTACACC	177
DB	1	AAGAACTTCCTCAGCTTTTGACATCCTGGTGGTGNCTGTGAGGACCTCGGCATGTATACAG	60
QY	178	TGCAGCGTGAGCAATGCGCTGGGACAGTGACACACCGGCGCTCTCCGGAAAGGCAGAG	237
DB	61	TGCAGTGTGAGCAATGCGCTGGGACGCGGCCACTACGCCATCTCTCCGAAGGCAGAG	120
QY	238	CGCCCTCATCTTCGCCATCCCGATATCGGGAGGTGTACGGGATGGGTGCTGCTCTG	297
DB	121	CGCCCTCTCTTCCTCGCCACGCGCGGACATTTGGGAGGTATACGCGGACGGGGTGTCTCTG	180
QY	298	GTCTGGAAGCCCGTGGAACTCTACGGGCCCTGTGACCTACATTTGTGAGTGCAGCTAGAA	357
DB	181	GTCTGGAAGCCTGTGGAGTCTACGGCCCTGTGACGTACATTTGTGAGTGCAGCTGGA	240
QY	358	GGCGGACGTGGACCACTATGGGCTCCGACATCTTTGACTGTGCTGTCTACCTGACACAGCAAG	417
DB	241	GGCGGACGTGGACACGCTGGCTCGGACGCTCTTTCAGTCTGCTACCTCACAGGAAG	300
QY	418	CTCTCCGGGGTGGCACCTACACCTTCGGACCGGCATGTGTACGACGGCAGGATGGGT	477
DB	301	CTTTCTCGGGTGGGGCGTACACCTTCGACCGCCTGCGTCAAGAGGGCGGCATGGGC	360
QY	478	CCCTACAGCAGCCCTCGGAGCAAGTCTCTCTGGGAGGGCCACGCCACCTTGGCCTCTGAG	537
DB	361	CCCTACAGCAGCCCTCGGAGCAGTCTCTCTTGGAGGGCCACGCCCTTGGCCTCTGAA	420
QY	538	GAGAGAGCCAGG---GGCGGTCTAGCCCAACCCCTGCCCCACACAAAGACTTTCGCATTC	594
DB	421	GAGAGAGTGGTGCACGGGGCCCGGCCGGCCCTGCCCCAGCATGACAGCTTCGCTTC	480

Qy	595	CAGACACAGATCCAGAGGGCCGCTT	CAGCGTGTGGGCAATGCTGGAGAGGCCAGC	655
Db	481	CAGACGAGATGAGAAGGGCCGCTT	CAGCGTGTGGGCAATGCTGGAGAGGCCAGC	540
Qy	655	GGCGGGCGCTGGCCG	670	
Db	541	GGCGGCATGCTGGCTG	556	
RESULT 7				
CR759117				
LOCUS		581 bp	mRNA	linear
DEFINITION		CR759117 mdx substracted from control	Mus musculus	cDNA clone
ACCESSION		POAB004ZB10		mRNA sequence.
VERSION		CR759117		
KEYWORDS		CR759117.1	GI:51879601	
SOURCE		EST.		
ORGANISM		Mus musculus	(house mouse)	
REFERENCE		1	(bases 1 to 581)	
AUTHORS		Rouger, K., LeCunff, M., Steenman, M., Potier, M.C., Gibelin, N., Dechesne, C.A. and Leger, J.J.		
TITLE		Global/temporal gene expression in diaphragm and hindlimb muscles of dystrophin-deficient (mdx) mice		
JOURNAL		Am. J. Physiol. Cell Physiol.	283 (C)	773-784 (2002)
COMMENT		Contact: Genoscope		
FEATURES		Genoscope - Centre National de Sequencage		
source		2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE		
		Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr		
		Genoscope sequence ID: POAB004ZB10F1.		
		Location/Qualifiers		
		1. 581		
		/organism="Mus musculus"		
		/mol_type="mRNA"		
		/db_xref="taxon:10090"		
		/clone="POAB004ZB10"		
		/sex="male"		
		/tissue_type="muscle"		
		/clone_lib="mdx substracted from control"		
		/notes="Vector: pCR11-TOPO; Pieu G., Cros N., Leger J.J., dechesne C.A. Mus musculus substracted libraries, detail of cDNA collection: pool of 4 cDNA libraries: C57BL/10mdx diaphragm cDNA substracted from C57BL/10ScSn diaphragm cDNA, C57BL/10ScSn diaphragm cDNA substracted from C57BL/10mdx diaphragm cDNA, C57BL/10mdx hindlimb muscle cDNA substracted from C57BL/10ScSn hindlimb muscle cDNA, C57BL/10ScSn hindlimb muscle cDNA substracted from C57BL/10mdx hindlimb muscle cDNA."		
ORIGIN				
Query Match		40.5%	Score 405.2;	DB 7; Length 581;
Best Local Similarity		82.5%;	Pred. No. 1.7e-81;	
Matches 477;		Conservative 0;	Mismatches 98;	Indels 3; Gaps 1
Qy	1	GGCCAGTCAGTGACACTGGCGCTG	CGCCAGGTCGACGCCAGCGAGTGC	CGCCAGGCCACCTGG 60
Db	4	GGCCATCATGTGACACTGGCTG	CGCCAGGTCGTCGCCAGCCACCACTGCCAGGCTACCTGG 63	
Qy	61	AGCAAGACGGAGCCCCCTCTGGAGAG	CAGACGCGTGTCTCATCTCTGCCACCTCAAG 120	
Db	64	AGCAAGATGGGTCTCTCTGGAGAG	CAGCGGCCACCTCTCATCTCTCCACCTGAAG 123	
Qy	121	AACCTCCAGCTTCTGACCATCTCTGGT	TGGTGGCTGAGGACCTGGGTGTGTACACCTGC 180	
Db	124	AACCTCCAGCTGCTGACCATCTCTGGT	TGGTGGTGAAGAGGAGGATCTGGGCACATATACCTGC 183	
Qy	181	AGCGTGAGCAATGCGCTGGGAGAC	GAGTGACACACACGGCGCTCTCCGGAAGCAGAGCGC 240	
Db	184	TGTGTAGCAACCCGCTGGGAGAC	GAGTACACACAGGTGCTCTCCGGAAGCAGAGCGC 243	
Qy	241	CCCTCATCTTTCCGCATCCCGGAT	TATCGGGAGGTGTACGCGGATGGGGTGTCTGCTGTC 300	

```

Db      244  CTTTCATCTTCTCCAGCCCTGAGTGGGGAACCTATACAAGGATCGGTGCTGCTGTC 303
QY      301  TGAAGCCCGTGAATCTTACGSCCTGTGACCTACATTGTGACGTGCGACCTAGAGGC 360
Db      304  TAGAAGCCGTGAATCTTGTGSCCGGTGACCTACATTGTGACGTGCTGTATAGAAGCA 363
QY      361  GGAGCTGGACCACTGSCCTCCGACATCTTTGACTGCTGTACTGACCAAGCAAGCTC 420
Db      364  GGAGCTGGACCACTGSCCTCCGACATCTTGTGACTGCTGTACTGACCAAGCAAGCTC 423
QY      421  TCCCGGGTGGACCTTACACTTCCGACGGATGTGTGAGCAAGCAAGCAAGCTTCC 480
Db      424  TCGAGGGTGGATGATATCTTCCGACAGCATGTGTGAGCAAGCAAGCAAGCTTCC 483
QY      481  TACAGCAGCCCTCGGAGCAAGTCTCTGAGGAGGCGCCAGCAAGCTTCTGAGGAG 540
Db      484  TACAGCAGTCCCTCAGACAGGTCTTCTAGGAGGAGCCCAACCACTGSCCTCTGAGGAG 543
QY      541  GAGAG---CGAGGGCGGTGAGCCCAACCCCTGCCCCAG 575
Db      544  GAAAGCAACCGGGAGGCGCCAGCTTTTTCCTCCAG 581

```

RESULT 8

CK601206
LOCUS
DEFINITION AGENCOURT 17900829 NIH_MGC_234 Rattus norvegicus cDNA clone
IMAGE:7190569 5', mRNA sequence.

ACCESSION CK601206
VERSION
KEYWORDS
SOURCE

ORGANISM Rattus norvegicus (Norway rat)

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1. (bases 1 to 771)

NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

CONTACT: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabbs-remail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LUAM15048 row: d column: 23

High quality sequence stop: 698.

Location/Qualifiers

1. 771

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="IMAGE:7190569"

/tissue_type="heart, pooled"

/lab_host="DH10B Tona"

/clone_lib="NIH_MGC_234"

/note="Organ: heart; Vector: pExpress-1; Site:1: EcoRV;

Site 2: NotI; RNA obtained from pooled heart tissue from a

mix of male and female animals at 8 wk old. Tissues were

snap-frozen and kept at -80C for two days before RNA

extraction and purification (TRI-reagent method). cDNA was

primed using oligo-dT primer:

5'-pGACTGTTCATGCTGCGAGCGCGCCCTT-25-3' and cloned into

the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb

resulted in an average insert size of 2.2 kb. This primary

library is normalized (non-normalized primary library is
NIH_MGC_233) and was constructed by Express Genomics
(Frederick, MD). Note: this is a NIH_MGC library."

ORIGIN

Query Match 39.7%; Score 397.4; DB 7; Length 771;
Best Local Similarity 82.1%; Pred. No. 1.1e-79;
Matches 469; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

```

QY      1  GGCAGTCAGTGACACTGGCCCTGCCAGTGTACGCCACGACGCTGCCAGGCACCTGG 60
Db      172  GGCCAAATCAGTGACACTGGCCCTGCCAGTGTGGCCAGCAACTGCCAGGCTACCTGG 231
QY      61  AGCAAGACGAGAGCCGCCCTGGAGAGCAGACCGGTGCTCTCATCTCTGCCACCTCAAG 120
Db      232  AGCAAGATGGGGCCCTTCTGGAGAGCAGCGCCACCTCTCTCATCTCTTCCACCTGAAG 291
QY      121  AACTTCCAGCTTCTGACCATCTCTGGTGTGGTCTGAGGACCTGGGTGTGTACACCTGC 180
Db      292  AACTTCCAGCTTCTGACCATCTCTGGTGTGGTGTGGGAGGATCTGGGCACATATACCTGC 351
QY      181  AGCTGTAGCAATGGCTGGGAGCAGTACCAACGAGGCGTCTCTCCGAAAGGAGAGGCG 240
Db      352  TGTGTGAGCAACCCACTAGGAGCAGCAGTACCACAGAGTGTCTCCGAAAGGAGAGGCG 411
QY      241  CCTTCATCTTCGCGCATGCCCGGATATCGGGAGGTGTACGGGATGGGTGCTGCTGTC 300
Db      412  CCTTCATCTTCTCACGCCCGGAGTGGGAACTATACAGGATGAGTGTGCTGCTGTC 471
QY      301  TGAAGCCCGTGGAAATCCTACGCGCTGTGACCTTACATTTGTCAGTGCAGCTAGAAAGGC 360
Db      472  TGAAGCCCTGTGGAATCCTATGCGCCAGTGACCTTACATTTGTCAGTGTCTGTATAGAAGCA 531
QY      361  GGAGCTGGACCACTGGGCTCGGACATCTTTGACGTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db      532  GGAGCTGGACCAACCCCTGGGCTCTGACATCTCCGACTGCTGCTGCTGCTGCTGCTGCTG 591
QY      421  TCCCGGGTGGCACCTTACATCTTCGACCGCATGTGTGACGAGGAGGAGCAAGTGGGTCCC 480
Db      592  CCGCGGGTGGCATGTATATCTTCGGACAGATGTGTGACGAAAGCAGGAGATGGGCCCC 651
QY      481  TACAGAGCCCTCGGAGCAAGTCTCTCTGGAGGCGCCAGCCACCTGGCTCTGAGGAG 540
Db      652  TACAGTAGCCCTCAGAAAGTCTCTCTGGAGGAGCCCA-CCACCTGGCTTCTGAGGAG 710
QY      541  GAGAGCCAGGGCGGTGAGCCCAACCCCTGC 571
Db      711  GAAGCAGCGGGAGAGCAGCCAGCTTCTCC 741

```

RESULT 9

CR758263/c

LOCUS

DEFINITION

CR758263 mdx substrated from control Mus musculus cDNA clone

POAB0042B10, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CR758263 565 bp mRNA linear EST 03-SEP-2004
CR758263 mdx substrated from control Mus musculus cDNA clone
POAB0042B10, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CR758263.1 GI:51878747
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 565)
Rouger, K., LeCunff, M., Steenman, M., Potier, M.C., Gibelin, N.,
Dechesne, C.A. and Leger, J.J.
Global/temporal gene expression in diaphragm and hindlimb muscles
of dystrophin-deficient (mdx) mice
Am. J. Physiol. Cell Physiol. 283 (C), 773-784 (2002)
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqrefgenoscope.cns.fr, Web: www.genoscope.cns.fr
Genoscope sequence ID: POAB0042B10R1.

FEATURES
source

Location/Qualifiers

1..565
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="POAB0042B10"
/sex="male"
/tissue_type="muscle"
/clone_lib="mdx substracted from control"
Dechesne C.A. Mus musculus substracted libraries, detail
of cDNA collection: pool of 4 cDNA libraries: C57BL/10mdx
diaphragm cDNA substracted from C57BL/10ScSn diaphragm
cDNA, C57BL/10ScSn diaphragm cDNA substracted from
C57BL/10mdx diaphragm cDNA, C57BL/10mdx hindlimb muscle
cDNA substracted from C57BL/10ScSn hindlimb muscle cDNA,
C57BL/10ScSn hindlimb muscle cDNA substracted from
C57BL/10mdx hindlimb muscle cDNA."

ORIGIN

Query Match 39.1%; Score 391.6; DB 7; Length 565;
Best Local Similarity 82.6%; Pred. No. 2.1e-78;
Matches 461; Conservative 0; Mismatches 94; Indels 3; Gaps 1;
QY 158 AGGACCTGGTGTACACCTGCAGCGTGAGCAATCGCTGGGACAGTGACACACGG 217
DB 558 AGGATCTGGGACATATACCTGCTGTGTAGCAACCCGCTGGGGACAGCGTCACACAG 499
QY 218 GCGTCTCCGGAAGCAGACAGCCCTCATCTTCGCCATGCCCCGATATCGGGAGGTGT 277
DB 498 GTGTCTCCGGAAGCAGACGCGCTTCATCTTCCAGCGCTCAGGTGGGGAACTAT 439
QY 278 AGCGGATGGGTGTCTGTGTCTGGAAGCCGCTGGAATCTTACCGCCCTGTGACCTACA 337
DB 438 ACAAGGATCGGTGTCTGTGTCTAGAAAGCCCTGTGGAATCTTGTGCGCGGTGACCTACA 379
QY 338 TTGTGACGTGACCTAGAAAGCGGACGTGACACACACTGCGCTCCGACATCTTTGACT 397
DB 378 TTGTGACGTGCTATAGAGTAGGACGTGGACAACTGCGCTTCAGACATCTCTGACT 319
QY 398 GCTGTACTGACCAAGCAAGCTCTCCCGGGTGGCACTTACACCTTCGCGACGGCATGTG 457
DB 318 GCTGTACTCTGCAAGCTATCGAGGGTGGCATGTATATCTTCGGACAGCATGTG 259
QY 458 TCAGCAAGCAGGAATGGTTCCTTACAGACGCGCTTCGGAGCAAGTCTCTCTGGAGGGC 517
DB 258 TCAGCAAGCAGGAATGGGCGCTTACAGACGCGCTTCAGAAAGTCTCTCTCGAGGAC 199
QY 518 CCAGCCACTGCGCTCTGAGGAGGA--GAGCGAGGGCGGTTCAGCCCAACCCCTGCCCA 574
DB 198 CCAACCACTGCGCTCTGAGGAGGAAGCAGCGGGGGAGGCGACGCCAGCTTCTCCCA 139
QY 575 GCACAAAGACCTTCGCTTCCAGACACAGATCCAGAGGGCGCTTCAGCGTGGTGGCGC 634
DB 138 GCACAAAGACCTTTGCTTCCAGATGCAGATCCGAGGGCGCTTCAGTGGTGGCGC 79
QY 635 AATGCTGGAGAGGCCAGCGCGCTGGCGCCCAAGATCATCCCTTACACCCCA 694
DB 78 AGTGAGGAGAGGAAGTGCGGGCGCTAGCTGCTAAGATCGTTCCTTACCAACCTG 19
QY 695 AGGCAAGACAGCACTGC 712
DB 18 AGGCAAGACAGCTGAC 1

RESULT 10

LOCUS BF134040 768 bp mRNA linear EST 24-OCT-2000
DEFINITION 601778492F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4006702 5',
mRNA sequence.
ACCESSION BF134040
VERSION BF134040.1 GI:10973080
KEYWORDS EST.

SOURCE
ORGANISM

Mus musculus (house mouse)

Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 768)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9238 row: g column: 23
High quality sequence stop: 736.

FEATURES

source

Location/Qualifiers

1..768

/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:4006702"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/clone_lib="NCI CGAP Lu30"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; transgenic model WNT-1, expression driven by
MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"

ORIGIN

Query Match 37.4%; Score 374.2; DB 2; Length 768;
Best Local Similarity 84.4%; Pred. No. 2e-74; Mismatches 445; Conservative 0; Indels 4; Gaps 2;
QY 478 CCCTACAGCAGCCCTCGGAGCAAGTCTCTCGGAGGGCCGAGCCACTGCGCTCTGAG 537
DB 2 CCCTACAGCAGCCCTCAGACAGGTCTCTCGGAGGAGCCACCACTTGGCTTGG 61
QY 538 GAGGA---GAGCAGGGGCGGTGAGCCCAACCCCTGCGCCAGCACAAAGACCTTCGATTTC 594
DB 62 GAGGAAAGCAGCGGGGAGGCGGAGCCAGCCAGCTTCTCCCGAGCACAAAGACCTTTCCTTC 121
QY 595 CAGACACAGATCCAGAGGGCGGCTTCAGCGTGGTGGCGCAATGCTGGGAGAGGCCAGC 654
DB 122 CAGATGCAGATCCGGAGGGGCGCTTCAGTGTGGTGGCGAGTGCAGGGAGAGCAAGT 181
QY 655 GGGCGGGCGCTGGCGCGCCCAAGATCATCCCTACCAACCCCAAGGACCAAGACAGCAGTGTG 714
DB 182 GGGCGGGCGCTAGCTGTAGATCGTTCCCTACCACTGAGGACCAAGACAGTGTACTA 241
QY 715 CGCGAATACAGGGCCCTCAAGGGCGTGGCCACCGCACCTGCGCCAGCTGCACAGCC 774
DB 242 AGAGAATACGAGGCACTTAAGAGACTGCACCAACCCACATCTGGCCCAACTCCATCGCGCC 301
QY 775 TACCTCAGCCCCCGGACCTGGTGTCTATCTTGGAGTGTGCTCTGGGCCCGAGTGTCTC 834
DB 302 TACCTCAGTCCCGGACCTGGTGTCTATCTTGGAGTGTGCTCTGGGCCCTGAGTGTCTA 361
QY 835 CCCTGCGCTGGCGGAGAGGGCGCTCTACTCAGAAATCCGAGGTGAAGGACTTACTCTGGGAG 894
DB 362 CCCTCTTTGGCGGAGAGGGGAATCTCTACAGATCTGATGTGAAGGACTTACTCTGGGAG 421
QY 895 ATGTTGAGTGCACCCAGTACCTGCACAAACAGACATCTCTGCACCTTGGACCTGAGGTCC 954
DB 422 ATGCTGAGTGCACCCAGTACTTGCATGCCCAACACATCTCTGCACCTTGGACCTGAGGTCC 481
QY 955 GAGAACATGATCATCACCAGCAATACACCTGCTCAAGGTCTGGGACCT 1001

[illegible]

Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

```

ORIGIN
Query Match      34.0%; Score 340.8; DB 4; Length 567;
Best Local Similarity 75.4%; Pred. No. 7.7e-67;
Matches 490; Conservative 0; Mismatches 77; Indels 83; Gaps 2;

QY 157 GAGGACCTGGGTGTACACCTGACGCTGAGCAATGCGTGGGACAGTAGACACCAAG 216
Db 1 GAGGACCTGGGCGTGTACACCTGACGCTGAGCAACGCACTGGGACAGCAACCAACA 60
QY 217 GCGCTCTCCGGAAGGAGAGCGCCCTCATCTTCGCCATGCGCCGATATGGGAGGTG 276
Db 61 GCCATCTCTCCAGAA-GCAGAGGCTCCCTCTCCCAAGCGCCGAGCATTTGGTGAATC 119
QY 277 TACGCGATGGGTGCTGCTGTGCTGGAAGCCGCTGGAATCTACGCGCCCTGTGACCTAC 336
Db 120 TATGCTGATGGGTGCTGCTGTGCTGGAAGCCCTGTGGAGTCATGTGGCCCTGTGACCTAT 179
QY 337 ATTGTGCAATGACGCTAGAGGGCGGAGCTGGACACACACTGGCTTCGACATCTTTGAC 396
Db 180 ATTGTGCAATGACGCTAGAGGGCGGAGCTGGAGCACATTTGGCCCTGACACATCTTTGAC 239
QY 397 TGCTGCTACCTGACGACCAAGCTCTCCCGGGGTGGCACCCTACACCTTCGCGACGGCATGT 456
Db 240 TGCTGCTACCTGACGCAAGCTTTCTCGGGGTGGGTATACACCTTCGCGACGGCCCTGT 299
QY 457 GTCAGCAAGGAGGAATGGGTCTCTACAGAGCCCTCGGACCAAGTCTCTCTGGGAGGG 516
Db 300 GTCAGCAAGGCGGATAGGCCCTCTACAGAGCCCTCGGAGCATCTCTCTGGGAGGA 359
QY 517 CCCAGCCACTTGGCCCTCTGAGGAGGAGCAGCGGGCGGTGAGCCCAACCCCTGCCACG 576
Db 360 CCCAGCCACTG----- 371
QY 577 ACAAGACCTTCGCATTCAGACACAGATCCAGAGGGCGGCTTCAGCGTGGTGGCGNA 636
Db 372 -----GGGCGCGCTTCAGTGTGGTGGAGGAG 397
QY 637 TGCTGGGAGAGGCCAGCGGGCGGCTGGCGGCAAGATCATCCCTACCAACCCCAAG 696
Db 398 TGCGGGAGAGGCCAGTGGGCGGCTGTGGTGCAGATGCTGCCCTACCAACCTGAG 457
QY 697 GACAAGACAGCATGTCTGCGGAATACAGAGCCCTCAAGGGCTGCGGACCCGCACTTG 756
Db 458 AACAGCACTGCGCTTCGAGATATGAGGCGCTCAAGGGCTTGGCGCACCCCTCACCTG 517
QY 757 GCCAGCTGCAGCAGCCTACTCTAGCCCGGCGACCTGGTCTCATCTT 806
Db 518 GCACAGTTGCAGGCTGCTTACCTCAGCCCGGCGACCTGGTCTCATCTT 567

```

```

RESULT 13
CF177813
LOCUS 806332 MARC 3P1G Sus scrofa cDNA 5', mRNA linear EST 28-JUL-2003
DEFINITION CF177813
ACCESSION CF177813
VERSION CF177813.1 GI:33289589
KEYWORDS EST.
SOURCE Sus scrofa (pig)

```

ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 494)
AUTHORS Smith,T.P.L., Fraking,B.A., Ford,J.J., Villet,J.L., Fox,J.,
Wise,T.A., Noneman,D.J., Wray,J.E. and Kaele,J.W.
TITLE A second set of porcine ESTs from a pooled-tissue normalized
library

JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.ueda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: SRG8004 row: M column: 6
Seq primer: GTATACGACTCACTATAGG.
Location/Qualifiers
1. 494

FEATURES

source
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3P1G"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN

```

Query Match      32.4%; Score 324.4; DB 7; Length 494;
Best Local Similarity 80.0%; Pred. No. 4.1e-63;
Matches 395; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

QY 505 CTCCTGGAGGGCCAGCCACCTGCGCTCTGAGGAGGAGAGCCAGG---GGCGGTACGCC 561
Db 1 CTCCTGGAGGACCCATGCGCTGCGCTCTGAAGAGGAGAGTGTGCTCATGACCGGCC 60
QY 562 CAACCCCTGCCAGCACAAAGACCTTCGCATTTCAGACACAGATCCAGAGGGGGCGCTTC 621
Db 61 CATCCCTCCCATCATCATACCTTTGCCCTTCCATACACAAATCATAAAGGGGGCGCTTC 120
QY 622 AGCGTGTGCGCAATGCTGGGAGAGGCCAGCGGGCGGCGCTGGCGGCCCAAGATCATC 681
Db 121 ATTGTGTGAGGCATTGCGGGAGNATGCCATTGGCGCGTGTGGCTGCCAATATCGTG 180
QY 682 CCCTACCAACCCCAAGGACAGACAGAGTGTGCGCGAATACGAGGCCCTCAAGGGCGCTG 741
Db 181 CGGGGCCACCCCTGAGAACAGGACTGCCGTGCTTCGATTAATATGATGCTCCCTCAAGGGCTG 240
QY 742 CGCCACCGCACCTGGCCCGAGCTGCAGCAGCCTACCTCAGCCCCCGGCACCTGGTGTCTC 801
Db 241 CGCCACCCCTACCTGGGCACATTTGCAGGCTGCCTACCTCCTCCCGGCGACCTGGTGTCTC 300
QY 802 ATCTTGGAGCTGTGCTCTGGCGCCGAGCTGTCTCCCTGCGCGAGAGGGCCCTCCTTAC 861
Db 301 ATCTTGGAGTTGCTCTGGCGCTGATCTGCTCCCTGCGCGAGAGGGCCCTCCTTAC 360
QY 862 TCAGATCCGAGGTGAAGGACTACCTGTGGCAGATGTTGAGTGCACCCAGTACCTGCAC 921
Db 361 TCATATACATATGTGAAGACTATCTGTGGCAGATGCTGATTTGCCACTCATTTACCTGCAC 420
QY 922 AACGACACATCTGCACCTGCAGCTGAGTCCGAGAACATGATCATCACCGAATACAC 981
Db 421 GCCAGCGCTATCTTCGATCTAGACCTCATGTCGGAACACATGATGTGCCCGAGTACAC 480
QY 982 CTGCTCAAGGTGCT 995
Db 481 CTGCTCAAGATCAT 494

```

```

RESULT 14
BY739930
LOCUS BY739930
DEFINITION RIKEN full-length enriched, pooled tissues, 16 days

```

embryo, etc. Mus musculus cDNA clone 1920063N21 5', mRNA sequence.
ACCESSION BY739930
VERSION BY739930.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

BY739930 653 bp mRNA linear EST 17-DEC-2002
BY739930 RIKEN full-length enriched, pooled tissues, 16 days
embryo, etc. Mus musculus cDNA clone 1920063N21 5', mRNA sequence.
ACCESSION BY739930
VERSION BY739930.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 653)

REFERENCE AUTHORS

Nikaido, I., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarilli, R., Hill, D.P., Butt, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Mutsaers, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.P., Forrest, A., Frazer, K.S., Gaasterland, I.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavalon, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Iehii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-Format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

Location/Qualifiers
1. .653
/organism="Mus musculus"
/mol_type="mRNA"

FEATURES source

/db_xref="taxon:10090"
/clone="I920063N21"
/clone_lib="RIKEN full-length enriched, pooled tissues, 16
days embryo, etc."
/note="pooled tissues: (dev stage=16 days
embryo, tissue type=heart, sex=mix), (dev stage=16 days
embryo, tissue type=kidney, sex=mix), (dev stage=17 days
embryo, tissue type=heart, sex=mix), (dev stage=17 days
embryo, tissue type=stomach, sex=mix), (dev stage=17 days
embryo, tissue type=kidney, sex=mix), (dev stage=17 days
pregnant, adult, tissue type=amion, sex=female),
(dev stage=13 days embryo, tissue type=liver, sex=mix)"

ORIGIN

Query Match 32.4%; Score 324.2; DB 6; Length 653;
Best Local Similarity 82.3%; Pred. No. 4.7e-63;
Matches 371; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
Qy 1 GCCCAGTCAGTGACACTGGCTGCGCAGGTGTGACGCCAGCGCTGCCAGGCCACCTGG 60
Db 203 GCCCAATCAGTGACACTGGCTGCGCAGGTGTGACGCCAGCGCTGCCAGGCCACCTGG 262
Qy 61 AGCAAAACGGAGCGCCCTGGAGAGCAGCAGCGGTCTCTCATCTCTGCCACCTCAAG 120
Db 263 AGCAAAAGATGGGTCTCTCTGGAGAGCAGCGGCCACCTCTCTCATCTCTGCCACCTGAAG 322
Qy 121 AACTTCCAGCTTCGACCATCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
Db 323 AACTTCCAGCTTCGACCATCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 382
Qy 181 AGCGTGAGCAATCGCTGGGGAGCAGTACCACCAACCGGGCTCTCCGGAAGGCAGAGCGC 240
Db 383 TGTGTGAGCAACCGCTGGGGAGCAGTACCACAGGTCTCTCCGGAAGGCAGAGCGC 442
Qy 241 CCCTCATCTTCCGCGATGCCCGGATATCGGGGAGGTGTACCGGAGTGGGTGGTGGTGGTGC 300
Db 443 CCTTCATCTTCTCCAGCCCTGAGTGGGGGAACTATACAAAGATGGGTGGTGGTGGTGC 502
Qy 301 TGGAAAGCCCTGGAATCTACGCCCTGTGACCTGTGACCTGTGACCTGTGACCTGTGACCTGTG 360
Db 503 TGGAAAGCCCTGGAATCTGTGCGCGGTGACCTGTGACCTGTGACCTGTGACCTGTGACCTGTG 562
Qy 361 GGAGCTGGAGCAACACTGGCTCCGACATCTTTGACTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 420
Db 563 GGAGCTGGAGCAACCTGGCTTCCAGACATCTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 622
Qy 421 TCCGGGGTGGCAGCTTACACCTTCCGACGG 451
Db 623 TCGAGGGGTGGCATGTATATCTTTCCGACAG 653

RESULT 15

CB713368
LOCUS
DEFINITION
AMGNNUC:NRHY7-00030-F2-A nrhy7 (10850) Rattus norvegicus cDNA clone
nrhy7-00030-f2 5', mRNA sequence.
CB713368
VERSION
CB713368.1 GI:29770516
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 494)
AUTHORS
Angen EST Program.
TITLE
Angen Rat EST Program
JOURNAL
Unpublished (2003)
COMMENT
Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00030 row: f column: 2.

THIS PAGE BLANK (USTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:02:07 ; Search time 70.3765 Seconds
(without alignments)
9323.380 Million cell updates/sec

Title: US-10-077-130-6_COPY_8500_8900
Perfect score: 401
Sequence: 1 gaagaccagtggtggcgcc.....caccagccaccagggagg 401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818130359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54.8	13.7	3397	4	US-09-949-016-2188
2	54.4	13.6	2768	4	Sequence 2188, Ap
3	51.2	12.8	7218	1	US-09-774-528-426
4	42.4	10.6	809	4	Sequence 14, Appl
5	42.4	10.6	822	4	US-09-949-016-2543
6	42.4	10.6	8260	4	Sequence 4657, Ap
7	42.4	10.6	14809	4	Sequence 14285, A
8	41.4	10.3	4403765	3	Sequence 1032, Ap
9	41.4	10.3	4411529	3	Sequence 2, Appl
10	41.2	10.3	789	3	US-09-103-840A-1
11	40.6	10.1	3575	4	Sequence 114, Appl
12	40.2	10.0	2873	4	Sequence 1395, Ap
13	40.2	10.0	289	3	Sequence 17, Appl
14	40	10.0	15644	4	Sequence 17, Appl
15	39	9.7	516	4	US-09-902-540-1133
16	39	9.7	6468	4	Sequence 1133, Ap
17	39	9.7	24281	4	Sequence 8502, Ap
18	38.6	9.6	1227	4	Sequence 895, Appl
19	38.6	9.6	2973	4	Sequence 14778, A
20	38.6	9.6	3595	4	Sequence 5523, Ap
21	38.6	9.6	9649	4	Sequence 4476, Ap
22	38.6	9.6	11254	4	Sequence 3658, Ap
23	38.6	9.6	23847	4	Sequence 15672, A
24	38.6	9.6	28172	4	Sequence 1040, Ap
25	38.4	9.6	1446	4	Sequence 1221, Ap
26	38.4	9.6	2022	4	Sequence 1177, Ap
27	38.4	9.6	10593	4	Sequence 3684, Ap
					Sequence 4403, Ap
					Sequence 1030, Ap

c 28	38.4	9.5	23847	4	US-09-902-540-1177	Sequence 1177, Ap
c 29	38	9.5	9080	4	US-09-902-540-1963	Sequence 1963, Ap
30	38	9.5	15209	4	US-09-902-540-1110	Sequence 1110, Ap
31	37.8	9.4	747	4	US-09-270-767-10494	Sequence 10494, A
32	37.6	9.4	729	4	US-09-902-540-4561	Sequence 4561, Ap
c 33	37.6	9.4	2409	4	US-09-902-540-6842	Sequence 6842, Ap
34	37.6	9.4	3899	4	US-09-902-540-573	Sequence 573, App
35	37.6	9.4	24905	4	US-09-902-540-1225	Sequence 1225, Ap
36	37.4	9.3	1707	4	US-09-949-016-2100	Sequence 2100, Ap
37	37.4	9.3	2032	4	US-09-902-540-8077	Sequence 8077, Ap
38	37.4	9.3	2424	4	US-09-902-540-4231	Sequence 4231, Ap
c 39	37.4	9.3	7811	4	US-09-902-540-824	Sequence 824, App
40	37.4	9.3	10644	4	US-09-902-540-1028	Sequence 1028, Ap
41	37.4	9.3	11465	4	US-09-949-016-13842	Sequence 13842, A
42	37.2	9.3	1905	4	US-09-902-540-2711	Sequence 2711, Ap
43	37.2	9.3	14861	4	US-09-902-540-1127	Sequence 1127, Ap
c 44	37.2	9.3	48974	3	US-08-920-422-17	Sequence 17, Appl
c 45	37	9.2	459	4	US-09-252-991A-9927	Sequence 9927, Ap

ALIGNMENTS

RESULT 1

US-09-949-016-2188
; Sequence 2188, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2188
; LENGTH: 3397
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2188

Query Match	13.7%	Score	54.8	DB	4	Length	3397
Best Local Similarity	47.2%	Pred. No.	0.00016				
Matches	167	Conservative	0	Mismatches	187	Indels	0
Gaps	0						
QY	12	GGTGGCGCCAGGGAGCGTGGAGCTGGCTGGTGGAGCTGTACCGGGCGGGAACCCCGT	71				
DB	354	GGTGAGCAAGGCCAACAGATCCCGCTGACCGTGAACCTGGTGACCTGACCTGAGGT	413				
QY	72	GCATCTGGCTGAAGGACAGAAAGGCATCCGCAAGAGCCAGAGTATGATGTGCTTGCGA	131				
DB	414	CAATGGCTCAAGATGTCAGAGGATCCAGATGAGCGGAGGATACATCTTTGAGTCCAT	473				
QY	132	GGGCAAGATGCCATGTGCTCATCCGGGGGCTTCGCTCAAGAGCGCGGGAGTACAC	191				
DB	474	CGGTGCCAAGCGTACCTGACCATCAGCCAGTGTCTCATTTGGCGGACGACGACCTACCA	533				
QY	192	GTGTGAGTGGAGGCTTCCAGAGCAGCAGCAGCTCCATGTGGAGAAAGCAACATG	251				
DB	534	GTGCGTGGTGGTGCGGAGAGTGTAGCAGCGAGCTCTTTGTGAAGAGCCCCCTGTGCT	593				
QY	252	CTTACAGAGGAGCTGACCAATCTGACGTGGAGAGAAAGCAGAGCTGTGTTCCAGTG	311				
DB	594	CATCAGCGCCCTTTGGAGGACCACTGTTGATGTTGGGGCAGCGGTTGAGTTGAGTG	653				
QY	312	CAAGCGGAGCACCCCGCGGCCACAGTACCTTGGCGCAAGGCGCTCTTTGGAGCT	365				

```
Db 654 TGAAGTATCGAGGAGGGGGCGCAAGTCAAAATGGCTGAAGACGGGTGGAGCT 707
|||||
RESULT 2
US-09-774-528-426
; Sequence 426, Application US/09774528
; Patent No. 6743619
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Kyle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Yang, Yonghong
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802
; CURRENT APPLICATION NUMBER: US/09/774,528
; CURRENT FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt_Fl_genes Version 2.0
; SEQ ID NO 426
; LENGTH: 2768
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72)..(2738)
US-09-774-528-426
Query Match 13.6%; Score 54.4; DB 4; Length 2768;
Best Local Similarity 48.9%; Pred. No. 0.00019;
Matches 179; Conservative 0; Mismatches 181; Indels 6; Gaps 1;

Qy 25 GAGGACGTGGAGCTCGCTGTGAGCTGTCAAGGCGGGAACCCCGTGCATGCTGTAAG 84
|||||
Db 2025 GAGGAGGCGCAGCTGCTGGCTGAGCTGTCAAGTCAAGCTGCGGCTGTGACGTGGCTGAAG 2084
|||||

Qy 85 GACAGGAAGGCATCCGCAAGAGCCAGCAAGTATGATGTGCTGCGAGGGCACGATGCC 144
|||||
Db 2085 GATGGTCGCACACTGTCCCCAGGCCCAAGTATGAGGTGCAGGCATCGCGCGGCGG 2144
|||||

Qy 145 ATGCTGGTTCATCCGCGGGCCCTCGCTCAAGGACGCGGCGAGTACACGTGTGAGTGGAG 204
|||||
Db 2145 GTGCTCTTGTGGAGATGTGCGCCGGGACGATGACGAGCCTCTACGATGGCTGAGCGCG 2204
|||||

Qy 205 GCTTCCAGAGCAGACCGCAGCTCCATGTGGAAGAAAGAAAAGTCTGTCCACAGAGGAG 264
|||||
Db 2205 GGGGCGGCATCGCTACCACTCTCCGTGCAAGGCTCGCGCTTTCTGCACAAGGAC 2264
|||||

Qy 265 CTG-----ACCAATCTCAGGTGGAGAGAAAGCACAGCTGTGTTACGTGCAAGACG 318
|||||
Db 2265 ATGCGGGCAGCTGTGTGGATGCCGTGGCTGGGGCGCGCGAGTTTGAAGTGTGAGACC 2324
|||||

Qy 319 GAGCACCCCGGGCCACAGTACCTTGGCGCAAGGCGCTTTGGAGCTACGGGCTCAGGG 378
|||||
Db 2325 TCCGAGCCACAGTCCACGTSCATGTGTACAAGGATGSCATGGAGCTGGGCCACTCCGGT 2384
|||||

Qy 379 AAGCAC 384
|||||
Db 2385 GAGCGC 2390
|||||

RESULT 3
```

```
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 12.8%; Score 51.2; DB 1; Length 7218;
Best Local Similarity 4.3%; Pred. No. 0.0015;
Matches 17; Conservative 219; Mismatches 162; Indels 0; Gaps 0;

Qy 1 GAAGACCAGTGGGTGGCGCCAGGGGAGAGCTGGAGCTGCGCTGTGAGCTGTCAAGGGCG 60
|||||
Db 1429 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1370
|||||

Qy 61 GGAACGCGCGTGCCTGCTGAGGACAGCAAGGCCATCCGCAAGAGCAGCAAGTATGAT 120
|||||
Db 1369 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1310
|||||

Qy 121 GTGCTGTGCGAGGCGACATGCGCATGTGTTATCGCGGGGCGCTGCTCAAGGACGCG 180
|||||
Db 1309 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1250
|||||

Qy 181 GCGGAGTACAGTGTGAGGTGGAGGCTTCCAAGAGCACAGCAGCTCCATGTGGAAGAA 240
|||||
Db 1249 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1190
|||||

Qy 241 AAAGCAAACTGCTTCACAGAGGAGCTGACCAATCTGAGGTGGAGGAGAAAGGACAGCT 300
|||||
Db 1189 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1130
|||||

Qy 301 GTGTTACGTGCAAGAGCGGAGCACCCCGCGCCACAGTGCACCTGGCGCAAGGGCTCTTG 360
|||||
```


Qy 193 TGTGAGGTGGAGCTTCAAGAGACACAGCCAGCCTCCATGTGGAGAAAAGCAAACTGC 252
Db 3414 GTCAATTGCTGCTCTCAAGGTCAAGTTCAAGCTCAAGGTTCATAGAGCGAGGTAAGATC 3473
Qy 253 TTCA 256
Db 3474 CTGA 3477

RESULT 7
US-09-902-540-1032
; Sequence 1032, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1032
; LENGTH: 14809
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1032

Query Match 10.6%; Score 42.4; DB 4; Length 14809;
Best Local Similarity 48.1%; Pred. No. 0.31;
Matches 151; Conservative 0; Mismatches 161; Indels 2; Gaps 1;
Qy 49 CTGTACGGGCGGGAACCCCGTGCATCTGAGGACAGGAGGCGCATCCGCAAGAGC 108
Db 2485 CTACGACGTGCGGAACGCC- -GGAGGGCGGCTGGGCTCTTACCCGCTCATCTC 2542
Qy 109 CAGAAGTATGATGTGTCGAGGGACGATGGCCATCTGGTTCATCCGGGGGCTCG 168
Db 2543 CACGCGAGGCGATGCGCGCGACGCGGCGCATGGACCTGGCCCTACGTGGCGCG 2602
Qy 169 CTCAGGACGCGGGGAGTACACGTGTGAGGTGAGGCTTCCAGAGACACAGCAGCCTC 228
Db 2603 GCGCGCTTCGAGCGCTTCTTGAGTTGAGCGGTGGACATCGCGCGGCTCG 2662
Qy 229 CATGTGGAAGAAAAGCAAACTGTTTCACAGAGAGCTGACCAATCTTCAGCGTGGAGAG 288
Db 2663 CTGCTGTGTGAGGAGGCGGGCGGTCTATCGTTCACATCTCCGGCGAGCCCTTCGACGTG 2722
Qy 289 AAAGGCACAGTGTGTTCACTGCAAGAGCGGACCCCGGCGCCACAGTGCATGGCGC 348
Db 2723 CTGCGCGCGACGTGATTGCTCCGCGCGCGCTGGCCCGCGGCTGCTGTCCAGGCG 2782
Qy 349 AAGGCGCTCTTGA 362
Db 2783 AAGCGCTTCGTGA 2796

RESULT 8
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 10.3%; Score 41.4; DB 3; Length 4403765;
Best Local Similarity 56.1%; Pred. No. 1.9;
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
Qy 28 GACGTGAGCTGCGCTGTGAGCTGTACGGCGGGAACGCCGTGCACTGGCTGAAGGAC 87
Db 1194752 GAGCGCGAGCGCAGCGGTCTGTTTACAGGGTGTGCGGCGCGACGACTTGTCTGACCGAA 1194693
Qy 88 AGGAAGCCATCCGCAAGAGCCAGAGTATGATGTGCTGCGAGGGCAGATGGCCATG 147
Db 1194692 GCCAGGCGCACTGCCACGACCATTTCCAGATGTGCGCTCGCGCGCCGATGCCCAAG 1194633
Qy 148 CTGTCATCCGCGGGGCT 166
Db 1194632 GAGCGCGTCAACCGGGCTT 1194614

RESULT 9
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 10.3%; Score 41.4; DB 3; Length 4411529;
Best Local Similarity 56.1%; Pred. No. 1.9;
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
Qy 28 GACGTGAGCTGCGCTGTGAGCTGTACGGCGGGAACGCCGTGCACTGGCTGAAGGAC 87
Db 1194537 GAGCGCGAGCGCAGCGGTCTGTTTACAGGGTGTGCGGCGCGACGACTTGTCTGACCGAA 1194478
Qy 88 AGGAAGCCATCCGCAAGAGCCAGAGTATGATGTGCTGCGAGGGCAGATGGCCATG 147
Db 1194477 GCCAGGCGCACTGCCACGACCATTTCCAGATGTGCGCTCGCGCGCCGATGCCCAAG 1194418
Qy 148 CTGTCATCCGCGGGGCT 166
Db 1194417 GAGCGCGTCAACCGGGCTT 1194399

RESULT 10
US-09-280-116-114/c
; Sequence 114, Application US/09280116A

Db	262	GGCGAGTTGCTGGCCGAGTTCAACACCGCCAGGACCCCGAGCGGCTCCACGTGGACGGA	321
Qy	241	AAAGCAAACCTGCTTCACACAGAGGAGCTGACCA	271
Db	322	GAACCATCTACCTGCTGGTGGAGCGCACCA	352

Search completed: March 21, 2005, 14:39:18
Job time : 85.3765 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:02:07 ; Search time 526.693 Seconds
(without alignments)
9323.380 Million cell updates/sec

Title: US-10-077-130-6_COPY_19500_22500
Perfect score: 3001
Sequence: 1 cctacggagcagaactcggg.....ccacctggagcaaacgga 3001

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2997.8	99.9	5207	4	US-09-858-664A-1
2	2997.8	99.9	5207	4	US-10-274-978-1
3	2997.8	99.9	5207	4	US-10-274-978-3
4	2997.8	99.9	5207	4	US-10-697-263-1
5	2997.8	99.9	5207	4	US-10-697-263-3
6	172	5.7	1429	2	US-09-159-385-4
7	172	5.7	1429	3	US-09-186-277-4
8	170.2	5.7	2157	4	US-09-949-016-1966
9	168.6	5.6	2105	4	US-09-949-016-195
10	168.6	5.6	2132	2	US-09-159-385-3
11	168.6	5.6	2132	3	US-09-186-277-3
12	148.6	5.0	1847	4	US-09-949-016-5449
13	148.6	5.0	5926	4	US-09-917-254-41
14	143	4.8	1788	4	US-09-797-039-9
15	143	4.8	2046	4	US-09-797-039-7
16	128.4	4.3	1584	4	US-09-799-451-205
17	122.2	4.1	1282	2	US-08-878-989-12
18	122.2	4.1	1282	3	US-09-272-796-12
19	122.2	4.1	1282	4	US-09-016-434-953
20	122.2	4.1	1349	1	US-07-951-715A-20
21	122.2	4.1	1349	2	US-08-459-448A-20
22	122.2	4.1	1349	3	US-08-459-595A-20
23	122.2	4.1	1349	3	US-08-459-504B-20
24	122.2	4.1	1349	3	US-08-459-444-20
25	122.2	4.1	1349	3	US-09-547-422-20
26	122.2	4.1	1349	4	US-09-988-462-20
27	121.2	4.0	8865	4	US-09-949-016-4238

28	121.2	4.0	8906	2	US-08-826-267-1	Sequence 1, Appli
29	118.6	4.0	1442	4	US-09-949-016-1590	Sequence 1590, Ap
30	118.6	4.0	1442	4	US-09-949-016-1591	Sequence 1591, Ap
31	118.6	4.0	1480	4	US-09-016-434-1454	Sequence 1454, Ap
32	116.2	3.9	1458	4	US-09-230-896C-5	Sequence 5, Appli
33	112.4	3.7	1074	4	US-09-733-388-3	Sequence 3, Appli
34	112.4	3.7	1074	4	US-10-446-175-3	Sequence 3, Appli
35	112.4	3.7	1158	4	US-09-733-388-1	Sequence 1, Appli
36	112.4	3.7	1158	4	US-10-446-175-1	Sequence 1, Appli
37	112.4	3.7	1671	4	US-09-733-388-5	Sequence 5, Appli
38	112.4	3.7	1671	4	US-10-446-175-5	Sequence 526, App
39	112.4	3.7	1733	4	US-09-620-312D-526	Sequence 3, Appli
40	101.4	3.4	1694	4	US-09-579-664B-3	Sequence 3, Appli
41	101.4	3.4	1694	4	US-10-355-975A-3	Sequence 1937, A
42	92	3.1	15376	4	US-09-949-016-11937	Sequence 13708, A
43	92	3.1	16621	4	US-09-949-016-13708	Sequence 15, Appl
44	90.4	3.0	5228	3	US-09-428-711A-15	Sequence 1, Appli
45	89.6	3.0	2447	4	US-09-960-643-1	

ALIGNMENTS

RESULT 1

US-09-858-664A-1

; Sequence 1, Application US/09858664A

; Patent No. 6482624

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui, et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL000927-CIP

; CURRENT APPLICATION NUMBER: US/09/858,664A

; CURRENT FILING DATE: 2001-05-17

; PRIOR APPLICATION NUMBER: 09/711,134

; PRIOR FILING DATE: 2000-11-11

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 5207

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-858-664A-1

Query Match 99.9%; Score 2997.8; DB 4; Length 5207;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	CCTACGGAGCAGAACTCGGGCCCGGATACAGGAGCGAGACATCTCTGGCCGCGCTGAG	60
Db	699	CCTACGGAGCAGAACTCGGGCCCGGATACAGGAGCGAGACATCTCTGGCCGCGCTGAG	758
Qy	61	CCACCCGCTGTCACCGGGGCTGTGGACCCGCAAGAGCCCTCATCTCTCAT	120
Db	759	CCACCCGCTGTCACCGGGGCTGTGGACCCGCAAGAGCCCTCATCTCTCAT	818
Qy	121	CCTGGAGCTGTGCTCATCCGAGAGCTGTGGACCGCTGTACAGGAAGCGGTGTGAC	180
Db	819	CCTGGAGCTGTGCTCATCCGAGAGCTGTGGACCGCTGTACAGGAAGCGGTGTGAC	878
Qy	181	GGAGCGCAGGTCAGGTCATCATCCAGAGCTGGTGGAGGGGTGCATCTACCTGCACAG	240
Db	879	GGAGCGCAGGTCAGGTCATCATCCAGAGCTGGTGGAGGGGTGCATCTACCTGCACAG	938
Qy	241	CCATGGGTTCTCCACCTGGACATAAAGCCCTCTTAACATCTCTGATGTGTCATCTGCCCG	300
Db	939	CCATGGGTTCTCCACCTGGACATAAAGCCCTCTTAACATCTCTGATGTGTCATCTGCCCG	998
Qy	301	GGAGACATTAATAATCTGCAGCTTTGGCTTTGCCAGAACATCAACCCAGAGAGTGCA	360
Db	999	GGAGACATTAATAATCTGCAGCTTTGGCTTTGCCAGAACATCAACCCAGAGAGTGCA	1058

Qy	361	GTTCAGCCAGTACGGCTCCCTGATGTTGCTCTCCCGGAGATCATTCAGACGAACCCCTGT	420
Db	1059	GTTCAGCCAGTACGGCTCCCTGATGTTGCTCTCCCGGAGATCATTCAGACGAACCCCTGT	1118
Qy	421	GAGCGAAGCCTCCGACATTTGGGCCATGGGTGTCATCTCTACCTCAGCTGACCTGCTC	480
Db	1119	GAGCGAAGCCTCCGACATTTGGGCCATGGGTGTCATCTCTCTACCTCAGCTGACCTGCTC	1178
Qy	481	ATCCCCATTTGGCCGGCAGAGTGACCGTGCCACCCCTCTGTAACGTCTCTGAGGGGGCCGT	540
Db	1179	ATCCCCATTTGGCCGGCAGAGTGACCGTGCCACCCCTCTGTAACGTCTCTGAGGGGGCCGT	1238
Qy	541	GTCATGAGCAGCCCATATGGCTGCCACCTCAGCGAAGACGCCAAAGACTTCATCAAGGC	600
Db	1239	GTCATGAGCAGCCCATATGGCTGCCACCTCAGCGAAGACGCCAAAGACTTCATCAAGGC	1298
Qy	601	TACGCTGCAGAGAGCCCTCAGGCGCCGGCTAGTGGCGGCCAGTGCCCTCCACCCCTG	660
Db	1299	TACGCTGCAGAGAGCCCTCAGGCGCCGGCTAGTGGCGGCCAGTGCCCTCCACCCCTG	1358
Qy	661	GTTCTCTGAAATTCATGCTCTCGGAGGAGGCCACTTCATCAACACCAAGCAGCTCAAGTT	720
Db	1359	GTTCTCTGAAATTCATGCTCTCGGAGGAGGCCACTTCATCAACACCAAGCAGCTCAAGTT	1418
Qy	721	CCTCTCGGCCGAAGTCGCTGGCAGCGTTTCCCTGATGAGCTACAAGTCCATCTCTGTGAT	780
Db	1419	CCTCTCGGCCGAAGTCGCTGGCAGCGTTTCCCTGATGAGCTACAAGTCCATCTCTGTGAT	1478
Qy	781	GGCTCCATCCCTGAGCTGTGCGGGGCCACCCGACAGCCCTCCCTCGGCTAGCCCG	840
Db	1479	GGCTCCATCCCTGAGCTGTGCGGGGCCACCCGACAGCCCTCCCTCGGCGTAGCCCG	1538
Qy	841	GCACCTCTGACGGAGACACTGGTGGCTCCTCCAGTTCTCTCTCTCTCTCTGACAAACAGCT	900
Db	1539	GCACCTCTGACGGAGACACTGGTGGCTCCTCTCAGTTCTCTCTCTCTCTCTGACAAACAGCT	1598
Qy	901	CGCCCCATTTGCCGGGCTAAGTCACTGCCACCCCTCCCGGTGACACACTCACCACTGCT	960
Db	1599	CGCCCCATTTGCCGGGCTAAGTCACTGCCACCCCTCCCGGTGACACACTCACCACTGCT	1658
Qy	961	GCACCCCGGGCTTCTTGCGGCGCTCGGCCAGCTGCTGAGGAAGCCGAGGCCAGTGA	1020
Db	1659	GCACCCCGGGGCTTCTTGCGGCGCTCGGCCAGCTGCTGAGGAAGCCGAGGCCAGTGA	1718
Qy	1021	GCCTCTCACCGAGGCCCCAGCTCGGCTGTCATCTCCGAGGGTCCGGGCCACCGGCCCG	1080
Db	1719	GCCTCTCACCGAGGCCCCAGCTCGGCTGTCATCTCCGAGGGTCCGGGCCACCGGCCCG	1778
Qy	1081	CAGGGGTGCTGCCCGGCACAGCTCATTCGACAGCTGTTCTTACCAAGCGGGTGA	1140
Db	1779	CAGGGGTGCTGCCCGGCACAGCTCATTCGACAGCTGTTCTTACCAAGCGGGTGA	1838
Qy	1141	GAGCCCTGAGCAGGGGCCCTGCCCCGGGAGCAGGCGCACCCGGCCGGCGGGCGCA	1200
Db	1839	GAGCCCTGAGCAGGGGCCCTGCCCCGGGAGCAGGCGCACCCGGCCGGCGGGCGCA	1898
Qy	1201	CCTGCTGAAGGGCGGTACATTTGCGGGGGCGCTGCCAGGCTCTGCGGAGCACAATGTA	1260
Db	1899	CCTGCTGAAGGGCGGTACATTTGCGGGGGCGCTGCCAGGCTCTGCGGAGCACAATGTA	1958
Qy	1261	GCACCGGGTCTGGAGGAGAGGCCGCCAGGAGGAGCAGGCCACCTCTGCGCCAAAGC	1320
Db	1959	GCACCGGGTCTGGAGGAGAGGCCGCCAGGAGGAGCAGGCCACCTCTGCGCCAAAGC	2018
Qy	1321	CCCTCATTCGAGACTGCCCTCGGCTGCTGCTCTGCAACCCCACTTGGCCCTGGCCA	1380
Db	2019	CCCTCATTCGAGACTGCCCTCGGCTGCTGCTCTGCAACCCCACTTGGCCCTGGCCA	2078
Qy	1381	CAGCCACTCCCTGGAACATGACTCTCCGAGCAGCCCGCCCTCTCTCGAGGCGCTGGG	1440
Db	2079	CAGCCACTCCCTGGAACATGACTCTCCGAGCAGCCCGCCCTCTCTCGAGGCGCTGGG	2138
Qy	1441	TGAGGCACAGCACTGCCCTTACGCCCTCTCCGGGGGGGGCCCCCTATTCAGGGACATGGGCA	1500

db	2139	 TGAGGCA	 CAGCGA	 CTGC	 CTTAC	 GCCCC	 CTCC	 GGGG	 GCCCC	 CTAT	 CAGGCA	2198
QY	1501	 CCCT	 CAGGGCT	 CTCA	 AGAG	 CTTC	 CACT	 TGCG	 CCAC	 CCAG	 CACT	1560
Db	2199	 CCCT	 CAGGGCT	 CTCA	 AGAG	 CTTC	 CACT	 TGCG	 CCAC	 CCAG	 CACT	2258
QY	1561	 GAGGC	 CAT	 CCCC	 GAG	 CGCT	 TGCG	 AGCG	 CCCC	 CTTC	 TGCG	1620
Db	2259	 GAGGC	 CAT	 CCCC	 GAG	 CGCT	 TGCG	 AGCG	 CCCC	 CTTC	 TGCG	2318
QY	1621	 TTCT	 TGCCCC	 CCCG	 GAG	 GGCT	 TGCG	 AGCG	 CCCC	 CTTC	 TGCG	1680
Db	2319	 TTCT	 TGCCCC	 CCCG	 GAG	 GGCT	 TGCG	 AGCG	 CCCC	 CTTC	 TGCG	2378
QY	1691	 CTT	 CCCT	 CAG	 GAT	 CTTT	 GCAA	 GAGG	 CCCC	 TTAG	 TACCT	1740
Db	2379	 CTT	 CCCT	 CAG	 GAT	 CTTT	 GCAA	 GAGG	 CCCC	 TTAG	 TACCT	2438
QY	1741	 GCCCC	 AGGCA	 CCCC	 CTGC	 CAAG	 CGAC	 CCCC	 CCCA	 TTGG	 ACCT	1800
Db	2439	 GCCCC	 AGGCA	 CCCC	 CTGC	 CAAG	 CGAC	 CCCC	 CCCA	 TTGG	 ACCT	2498
QY	1801	 TGGAG	 CAT	 CTCT	 CTT	 CTTG	 GGAG	 GGCC	 AAAA	 CCCC	 GGCT	1860
Db	2499	 TGGAG	 CAT	 CTCT	 CTT	 CTTG	 GGAG	 GGCC	 AAAA	 CCCC	 GGCT	2558
QY	1861	 CTCC	 AGG	 CAG	 CTT	 TCCC	 AAG	 TAG	 CTCC	 TCA	 GGT	1920
Db	2559	 CTCC	 AGG	 CAG	 CTT	 TCCC	 AAG	 TAG	 CTCC	 TCA	 GGT	2618
QY	1921	 AGAC	 CT	 TGG	 CCCC	 CTCT	 GTAG	 TCCG	 AGG	 CTGT	 GAG	1980
Db	2619	 AGAC	 CT	 TGG	 CCCC	 CTCT	 GTAG	 TCCG	 AGG	 CTGT	 GAG	2678
QY	1981	 CTCC	 ACA	 CCCT	 TGAC	 GGCC	 CTCA	 GAG	 CGAC	 CCAT	 TGCC	2040
Db	2679	 CTCC	 ACA	 CCCT	 TGAC	 GGCC	 CTCA	 GAG	 CGAC	 CCAT	 TGCC	2738
QY	2041	 TGCT	 CGG	 GGG	 CTAC	 GCA	 GGT	 TGCT	 GCA	 CTTG	 CGTAG	2100
Db	2739	 TGCT	 CGG	 GGG	 CTAC	 GCA	 GGT	 TGCT	 GCA	 CTTG	 CGTAG	2798
QY	2101	 AGGG	 GCA	 TCT	 CGGG	 CACA	 TG	 GGG	 CCCA	 TAG	 GCT	2160
Db	2799	 AGGG	 GCA	 TCT	 CGGG	 CACA	 TG	 GGG	 CCCA	 TAG	 GCT	2858
QY	2161	 GGAG	 GAG	 GAG	 GCG	 CCAG	 GGCT	 TAG	 TCCC	 AGT	 CCAG	2220
Db	2859	 GGAG	 GAG	 GAG	 GCG	 CCAG	 GGCT	 TAG	 TCCC	 AGT	 CCAG	2918
QY	2221	 CAGGG	 CTG	 AGCC	 CACT	 GCG	 CCCA	 TG	 GGG	 CCAG	 GAT	2280
Db	2919	 CAGGG	 CTG	 AGCC	 CACT	 GCG	 CCCA	 TG	 GGG	 CCAG	 GAT	2978
QY	2281	 TCCG	 ACG	 AGCT	 CTCG	 AGCC	 CAAC	 				
Db	2979	 TCCG	 ACG	 AGCT	 CTCG	 AGCC	 CAAC	 				
QY	2341	 GCAG	 T	 CCGG	 ACCT	 GATG	 TCGG	 AGCG	 CCG	 CCG	 CA	2400
Db	3039	 GCAG	 T	 CCGG	 ACCT	 GATG	 TCGG	 AGCG	 CCG	 CCG	 CA	3098
QY	2401	 CGAG	 TGB	 AC	 CCG	 CTTA	 CCT	 CA	 GCT	 CTCA	 GAT	2460
Db	3099	 CGAG	 TGB	 AC	 CCG	 CTTA	 CCT	 CA	 GCT	 CTCA	 GAT	3158
QY	2461	 CGAG	 T	 TAT	 GAT	 CTTC	 CAG	 AA	 AGT	 CCCC	 AAGT	2520
Db	3159	 CGAG	 T	 TAT	 GAT	 CTTC	 CAG	 AA	 AGT	 CCCC	 AAGT	3218
QY	2521	 GGCT	 TAG	 GAG	 GAG	 CT	 TCCG	 AGCC	 CA	 GCTG	 GGC	2580

```
Db 3219 GGCTGAGGAGAGCTGGCCGAGTTCCCGGAGCCACAGTGGCCCTGGCCAGGTGAACCTGGG 3278
Qy 2581 CCCCCACGCGGCTTGAGATCAAGAGGAGTCAAGAGGATGAGACGCGCTGCTGGCAGA 2640
Db 3279 CCCCCACGCGGCTTGAGATCAAGAGGAGTCAAGAGGATGAGACGCGCTGCTGGCAGA 3338
Qy 2641 GGCTGCGGTGGGACGAGGAGCGCAAGTGGTCTCGCCGTCAGCAGGCGCTCTTCCACTTCCC 2700
Db 3339 GGCTGCGGTGGGACGAGGAGCGCAAGTGGTCTCGCCGTCAGCAGGCGCTCTTCCACTTCCC 3398
Qy 2701 TGGGAGGACACCTGCGCGCTGAGTGCAGAGCTGGGGCTGGCGTGAGAGAGTGAAGGC 2760
Db 3399 TGGGAGGACACCTGCGCGCTGAGTGCAGAGCTGGGGCTGGCGTGAGAGAGTGAAGGC 3458
Qy 2761 CTCCGTGAGACATCTCCCGGATCTGAAAGGAGCGCCGAGAGGTCGAGAGAGGAGG 2820
Db 3459 CTCCGTGAGACATCTCCCGGATCTGAAAGGAGCGCCGAGAGGTCGAGAGAGGAGG 3518
Qy 2821 GCCCCCAGGAGAGGAGCCAGGCTTGTCTTCCGCGCTCAGGTCTGAAGAGCTGGGA 2880
Db 3519 GCCCCCAGGAGAGGAGCCAGGCTTGTCTTCCGCGCTCAGGTCTGAAGAGCTGGGA 3578
Qy 2881 CCAGAGCGCCACATCTCTAAGGAGCTCTCAGATGAGACTGTGTCTTGGGCCAGTCAGT 2940
Db 3579 CCAGAGCGCCACATCTCTAAGGAGCTCTCAGATGAGACTGTGTCTTGGGCCAGTCAGT 3638
Qy 2941 GACACTGCGCTGCGAGGTGTCAGCCAGCAGCTGCCCCAGCCACCTGGAGCAAGACGG 3000
Db 3639 GACACTGCGCTGCGAGGTGTCAGCCAGCAGCTGCCCCAGCCACCTGGAGCAAGACGG 3698
Qy 3001 A 3001
Db 3699 A 3699

RESULT 2
US-10-274-978-1
; Sequence 1, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
US-10-274-978-1

Query Match 99.9%; Score 2997.8; DB 4; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCTACGGAGCAGAACTCGGGCCAGGATACAGGAGCGAGACATCTTGGCCGCGCTGAG 60
Db 699 CCTACGGAGCAGAACTCGGGCCAGGATACAGGAGCGAGACATCTTGGCCGCGCTGAG 758
Qy 61 CCACCCGCTGTCAAGGGGCTGCTGGACCAAGTTTGAACCCGCAAGACCTCATCTCTAT 120
Db 759 CCACCCGCTGTCAAGGGGCTGCTGGACCAAGTTTGAACCCGCAAGACCTCATCTCTAT 818
Qy 121 CCTGGAGCTGTCTATCCGAGGAGCTGCTGGACCGCTGTACAGGAGGCGGTGGTAC 180
Db 121 CCTGGAGCTGTCTATCCGAGGAGCTGCTGGACCGCTGTACAGGAGGCGGTGGTAC 1958
```

```
Db 819 CCTGGAGCTGTCTATCCGAGGAGCTGCTGGACCGCCTGTACAGGAAGGCGGTGGTGAC 878
Qy 181 GGAGGCCGAGGTCAAGGTCTAATCCAGAGAGCTGGTGGAGGGGTGCACTACTTGCACAG 240
Db 879 GGAGGCCGAGGTCAAGGTCTAATCCAGAGAGCTGGTGGAGGGGTGCACTACTTGCACAG 938
Qy 241 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTAAACATCCTGATGGTGCATCTGCCCG 300
Db 939 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTAAACATCCTGATGGTGCATCTGCCCG 998
Qy 301 GGAAGACATTAATAATCTGCGACTTTTGGCTTTTGGCCAGAAACATCAACCCAGCAGAGTGA 360
Db 999 GGAAGACATTAATAATCTGCGACTTTTGGCTTTTGGCCAGAAACATCAACCCAGCAGAGTGA 1058
Qy 361 GTTCAGCAGTACGGCTCCCTGAGTTGCTCTCCCCGAGATCATCCAGCAGAAACCTGT 420
Db 1059 GTTCAGCAGTACGGCTCCCTGAGTTGCTCTCCCCGAGATCATCCAGCAGAAACCTGT 1118
Qy 421 GAGCGAAGCTCCGACATTTGGGCCATGGGTGTCTCTTACCTCAGCTCAGCTGCTGCTC 480
Db 1119 GAGCGAAGCTCCGACATTTGGGCCATGGGTGTCTCTTACCTCAGCTCAGCTGCTGCTC 1178
Qy 481 ATCCCCATTTGCCGCGCAGAGTGAACCGTGGCCACCTCTCTGAAACGCTCTGGAGGGCGCGT 540
Db 1179 ATCCCCATTTGCCGCGCAGAGTGAACCGTGGCCACCTCTCTGAAACGCTCTGGAGGGCGCGT 1238
Qy 541 GTCATGAGAGCGCCCATGCTGCCACCTCAGCGAAGACGCCAAAGACTTTCATCAAGGC 600
Db 1239 GTCATGAGAGCGCCCATGCTGCCACCTCAGCGAAGACGCCAAAGACTTTCATCAAGGC 1298
Qy 601 TAGCTGTCAGAGAGCCCTCAGGCGCGGCTAGTGGCGGCCAGTGCCTCTCCACCCCTG 660
Db 1299 TAGCTGTCAGAGAGCCCTCAGGCGCGGCTAGTGGCGGCCAGTGCCTCTCCACCCCTG 1358
Qy 661 GTTCCTGAAATCCATGCTCGGAGGAGGCGCCACTTCATCAACCAAGCAGAGCTCAAGTT 720
Db 1359 GTTCCTGAAATCCATGCTCGGAGGAGGCGCCACTTCATCAACCAAGCAGAGCTCAAGTT 1418
Qy 721 CTCTCTGGCCGAAAGTCTGCTGGCAGCGTTCCTCTGATGAGCTACAAGTTCATCTCTGAT 780
Db 1419 CTCTCTGGCCGAAAGTCTGCTGGCAGCGTTCCTCTGATGAGCTACAAGTTCATCTCTGAT 1478
Qy 781 GCGTCTCATCTCTGAGCTGCTGGGGGCCACCCAGCAGAGCCCTCCCTCGGCGGTAGCCCG 840
Db 1479 GCGTCTCATCTCTGAGCTGCTGGGGGCCACCCAGCAGAGCCCTCCCTCGGCGGTAGCCCG 1538
Qy 841 GCACCTCTGAGGAGACACTGCTGGCTCTCTCAGTTCTCTCTCTCTCTGACAAAGAGCT 900
Db 1539 GCACCTCTGAGGAGACACTGCTGGCTCTCTCAGTTCTCTCTCTCTCTGACAAAGAGCT 1598
Qy 901 CGCCCCATTTGCCCGGCTAAGTCACTGCGACCCCTCCCGGTGACACACTCACCAGTCT 960
Db 1599 CGCCCCATTTGCCCGGCTAAGTCACTGCGACCCCTCCCGGTGACACACTCACCAGTCT 1658
Qy 961 GCACCCCGGGGCTTCTCGGCGGCTCGGCGGCTGCTGAGGAAGCCGAGGCGCAGTGA 1020
Db 1659 GCACCCCGGGGCTTCTCGGCGGCTCGGCGGCTGCTGAGGAAGCCGAGGCGCAGTGA 1718
Qy 1021 GCGTCTCACCGAGGCGCCAGCTCCGCTGCTATCTCCGAGGGTCCCGGGCCACCCGGCGCGC 1080
Db 1719 GCGTCTCACCGAGGCGCCAGCTCCGCTGCTATCTCCGAGGGTCCCGGGCCACCCGGCGCGC 1778
Qy 1081 CCAGGGTGTGCTGCCCGGACAGCGTCACTCCGAGCCTGTTTCAACACAGGCGGGTGA 1140
Db 1779 CCAGGGTGTGCTGCCCGGACAGCGTCACTCCGAGCCTGTTTCAACACAGGCGGGTGA 1838
Qy 1141 GAGCCCTGAGCAAGGGGCTTGGCCCGGGGAGCAGCGGCGCACCCGGCCCGGGCGCGCA 1200
Db 1839 GAGCCCTGAGCAAGGGGCTTGGCCCGGGGAGCAGCGGCGCACCCGGCCCGGGCGCGCA 1898
Qy 1201 CTCTCTAAGGGGCGGTACATTGGGGGGCGCTGCCAGGCTGCGCGAGGCGCTGATGGA 1260
Db 1899 CTCTCTAAGGGGCGGTACATTGGGGGGCGCTGCCAGGCTGCGCGAGGCGCTGATGGA 1958
```


Query Match					99.9%; Score 2997.8; DB 4; Length 5207;				
Beat Local Similarity					99.9%; Pred. No. 0;				
Matches 2999; Conservative					0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1	CCTACGGAGCAGAACTCGGGCCCGAGGCAATACAGGAGGAGAGACATCTCTGGCGCGGCTGAG	60						
Db	699	CCTACGGAGCAGAACTCGGGCCCGAGGCAATACAGGAGGAGAGACATCTCTGGCGCGGCTGAG	758						
Qy	61	CCACCCGCTGTACGGGGCTGTGGACCAAGTTTGGACCCCGGAGACCTTCATCTCTCAT	120						
Db	759	CCACCCGCTGTGTACGGGGCTGTGGACCAAGTTTGGACCCCGGAGACCTTCATCTCTCAT	818						
Qy	121	CCTGGAGCTGTCTATCCGAGGAGCTGTGACCGCCTGTACAGGAAGGCGGTGTGAC	180						
Db	819	CCTGGAGCTGTCTATCCGAGGAGCTGTGACCGCCTGTACAGGAAGGCGGTGTGAC	878						
Qy	181	GGAGCCGAGGTCAAGGTCTACATCCAGCAGCTGTGGAGGGGTGCACTACCTGCAAG	240						
Db	879	GGAGCCGAGGTCAAGGTCTACATCCAGCAGCTGTGGAGGGGTGCACTACCTGCAAG	938						
Qy	241	CCATGGGCTTCTCACCTTGACATAAAGCCCTCTAAACATCTGTATGTGTGATCTGTGCCG	300						
Db	939	CCATGGGCTTCTCACCTTGACATAAAGCCCTCTAAACATCTGTATGTGTGATCTGTGCCG	998						
Qy	301	GGAGACATTAATAATCTGCGACTTTGGCTTTGCCAGAACATCAACCCAGCAGAGCTGCA	360						
Db	999	GGAGACATTAATAATCTGCGACTTTGGCTTTGCCAGAACATCAACCCAGCAGAGCTGCA	1058						
Qy	361	GTTCAGCCAGTACGGCTCCCTGAGTTCTGCTCTCCCGAGATCATCCAGCAGAACCTGT	420						
Db	1059	GTTCAGCCAGTACGGCTCCCTGAGTTCTGCTCTCCCGAGATCATCCAGCAGAACCTGT	1118						
Qy	421	GAGCGAGCCTCCGACATTTGGGCAATGGGTGTGATCTCTACCTCAGCTGACCTGTCTC	480						
Db	1119	GAGCGAAGCCTCCGACATTTGGGCAATGGGTGTGATCTCTACCTCAGCTGACCTGTCTC	1178						
Qy	481	ATCCCCATTTCCCGCGAGAGTACCGTGGCCACCTCTCTGAACCTCTCTGGAGGGCGCGT	540						
Db	1179	ATCCCCATTTCCCGCGAGAGTACCGTGGCCACCTCTCTGAACCTCTCTGGAGGGCGCGT	1238						
Qy	541	GTCTAGAGCAGCCCATGGCTGGCCACCTCAGCGAAGACGCCAAGACTTTCATCAAGGC	600						
Db	1239	GTCTAGAGCAGCCCATGGCTGGCCACCTCAGCGAAGACGCCAAGACTTTCATCAAGGC	1298						
Qy	601	TACGCTGAGAGAGCCCTCAGGCCCGGCTAGTGGGCCCAAGTGCCTCTCCCAACCCCTG	660						
Db	1299	TACGCTGAGAGAGCCCTCAGGCCCGGCTAGTGGGCCCAAGTGCCTCTCCCAACCCCTG	1358						
Qy	661	GTTCCTGAATCCATGCTCGGAGGAGGCCATCTCATCAACCAAGCAGCTCAAGTT	720						
Db	1359	GTTCCTGAATCCATGCTCGGAGGAGGCCATCTCATCAACCAAGCAGCTCAAGTT	1418						
Qy	721	CCTCTGGCCCGAAGTCTGGCAGCGTTCCCTGATGAGCTTACAGTCCATCTCTGGTGAT	780						
Db	1419	CCTCTGGCCCGAAGTCTGGCAGCGTTCCCTGATGAGCTTACAGTCCATCTCTGGTGAT	1478						
Qy	781	CGCTTCCATCTCCTGAGTGTCTGGGGCCCAACCCGACAGCCCTCTCCCTCGGCTAGCCCG	840						
Db	1479	CGCTTCCATCTCCTGAGTGTCTGGGGCCCAACCCGACAGCCCTCTCCCTCGGCTAGCCCG	1538						
Qy	841	GCACCTCTGAGGAGACACTGTGTGCTCTCCAGTTCTCTCTCTCTCTGACAGAGCT	900						
Db	1539	GCACCTCTGAGGAGACACTGTGTGCTCTCCAGTTCTCTCTCTCTCTGACAGAGCT	1598						
Qy	901	CGCCCAATTTGCCCGGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACCACTGCT	960						
Db	1599	CGCCCAATTTGCCCGGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACCACTGCT	1658						
Qy	961	GCACCCCGGGCTTCTCTGGGCCCTCTGGCCAGCTCTGCTGAGGAAGCCGAGGCCAGTGA	1020						
Db	1659	GCACCCCGGGCTTCTCTGGGCCCTCTGGCCAGCTCTGCTGAGGAAGCCGAGGCCAGTGA	1718						
Qy	1021	CGGCTCCACCGAGGCCCTCCGCTCTGATCTCCCGAGGGTGTCCGGGCCACCGGCCG	1080						

Db	1719	GGCTTCCACCGAGGCCCCAGCTCCGCTTGCATCTCCCGAGGGTGTCCGGGCCACCGCGCGC	1778						
Qy	1081	CCAGGGTGTGCTGCTCCCGGCAACAGCGTCACTCCGAGGCTGTCTTACCAACAGGGGGTGA	1140						
Db	1779	CCAGGGTGTGCTGCTCCCGGCAACAGCGTCACTCCGAGGCTGTCTTACCAACAGGGGGTGA	1838						
Qy	1141	GAGCCCTGAGCAACGGGGCTTGGCCCCGGGGAGCAGGGGCAACCGGCCCGCGCGCGCA	1200						
Db	1839	GAGCCCTGAGCAACGGGGCTTGGCCCCGGGGAGCAGGGGCAACCGGCCCGCGCGCGCA	1898						
Qy	1201	CCTCTCAAGGGGGGTACATTGGGGGGGCTGCCAGGCTGTGCGAGCCACTGTATGGA	1260						
Db	1899	CCTCTCAAGGGGGGTACATTGGGGGGGCTGCCAGGCTGTGCGAGCCACTGTATGGA	1958						
Qy	1261	GCACCGCTGTGAGGAGGAGCGCCAGGGAGGAGCAGGCCACCTCTCTGGCCAAAGC	1320						
Db	1959	GCACCGCTGTGAGGAGGAGCGCCAGGGAGGAGCAGGCCACCTCTCTGGCCAAAGC	2018						
Qy	1321	CCCTCATTCGAGACTGCCCTTCGGGCTGCCTCTGCGACCCACTTGGCCCCCTGGCCA	1380						
Db	2019	CCCTCATTCGAGACTGCCCTTCGGGCTGCCTCTGCGACCCACTTGGCCCCCTGGCCA	2078						
Qy	1381	CAGCACTCTCTGGAACATGACTCTCGAGCACCCCGCCCTCTCTCGAGGGCTTGGG	1440						
Db	2079	CAGCACTCTCTGGAACATGACTCTCGAGCACCCCGCCCTCTCTCGAGGGCTTGGG	2138						
Qy	1441	TGAGGCAACAGCACTGCTTTCAGCCCCCTTCGGGGGGGGCCCCCTATCAGGGACATGGGGCA	1500						
Db	2139	TGAGGCAACAGCACTGCTTTCAGCCCCCTTCGGGGGGGGCCCCCTATCAGGGACATGGGGCA	2198						
Qy	1501	CCCTCAGGGCTCCAGCAGCTTCATCCTGCTGGGCAACCCAGGCACTGCTCAGGCCAGA	1560						
Db	2199	CCCTCAGGGCTCCAGCAGCTTCATCCTGCTGGGCAACCCAGGCACTGCTCAGGCCAGA	2258						
Qy	1561	GAGGCATCTCCCGACAGCCTTGGGGGAGCAGCCAGCCCTTTCTGCGACCCCAAGCAGGG	1620						
Db	2259	GAGGCATCTCCCGACAGCCTTGGGGGAGCAGCCAGCCCTTTCTGCGACCCCAAGCAGGG	2318						
Qy	1621	TTCTGCCCCCAGAGGGGTGACGCCCCCAACCCAGCAGTTTGCCTCATTCCTCTGGGCTC	1680						
Db	2319	TTCTGCCCCCAGAGGGGTGACGCCCCCAACCCAGCAGTTTGCCTCATTCCTCTGGGCTC	2378						
Qy	1681	CTTCCCTTCAGGATCTTTGCAAGAGGGCCCCCTTAGTACCTCTCAAGCCCTCTTCTGGGCA	1740						
Db	2379	CTTCCCTTCAGGATCTTTGCAAGAGGGCCCCCTTAGTACCTCTCAAGCCCTCTTCTGGGCA	2438						
Qy	1741	GCCCCAGGCACCCCTCGCCCTGCAAGCAAGCCCCCACTTGGACTCTAAGATGGGGCC	1800						
Db	2439	GCCCCAGGCACCCCTCGCCCTGCAAGCAAGCCCCCACTTGGACTCTAAGATGGGGCC	2498						
Qy	1801	TGGAGACATCTCTTCTTCTGGAGGCCAAAACCCGGGCCCTTGCAAGTTCCCCAGGGTTCAGC	1860						
Db	2499	TGGAGACATCTCTTCTTCTGGAGGCCAAAACCCGGGCCCTTGCAAGTTCCCCAGGGTTCAGC	2558						
Qy	1861	CTCCAGGAGGAGCTCTTCCCAAGTGAAGCTCCTCAGGGTGGGCTCTCTCCAGGTGGGCAC	1920						
Db	2559	CTCCAGGAGGAGCTCTTCCCAAGTGAAGCTCCTCAGGGTGGGCTCTCTCCAGGTGGGCAC	2618						
Qy	1921	AGAGCTGGGCCCTCTCCCTGGATGCGAGGGCTGGACCCAGAGGCTGAGGATCTGTCCGA	1980						
Db	2619	AGAGCTGGGCCCTCTCCCTGGATGCGAGGGCTGGACCCAGAGGCTGAGGATCTGTCCGA	2678						
Qy	1981	CTCCACACCCACCTTTCAGCGGGCTCTAGGAACAGGCGACCTAGCGCAAGTTCTCCCTGGG	2040						
Db	2679	CTCCACACCCACCTTTCAGCGGGCTCTAGGAACAGGCGACCTAGCGCAAGTTCTCCCTGGG	2738						
Qy	2041	TGGTTCGGGGGGGTACCGCAGGCTGTGGCTATGCGACCTTTGCTCTTGGTGAGATGC	2100						
Db	2739	TGGTTCGGGGGGGTACCGCAGGCTGTGGCTATGCGACCTTTGCTCTTGGTGAGATGC	2798						
Qy	2101	AGGGGGACATCTGGGGGAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGCTCCAGTC	2160						

Db 2799 AGGGGGCATGCTGGGGCAGGGGCCCATGTTGGGCCAGGATAGCTGGGCTGTGTCCAGTC 2858
Qy 2161 GGAGGAGGAGCAGCAGAGGAGGCGCAGGGCTGTAGTCCAGTTCGAGGAGCAGCAGGAGGC 2220
Db 2859 GGAGGAGGAGCAGCAGAGGAGGCGCAGGGCTGTAGTCCAGTTCGAGGAGCAGCAGGAGGC 2318
Qy 2221 CAGGGCTGAGAGGCCACTGCCCCCAGGTCACTGCAAGGCCCTGTGCTGAGGTTCGGCAGGCG 2280
Db 2919 CAGGGCTGAGAGGCCACTGCCCCCAGGTCACTGCAAGGCCCTGTGCTGAGGTTCGGCAGGCG 2978
Qy 2281 TCCACACAGAGCTCTCCAGAGCCACCCCATGGGAGGACATCGGGAGGTCTCCCTGGT 2340
Db 2979 TCCACACAGAGCTCTCCAGAGCCACCCCATGGGAGGACATCGGGAGGTCTCCCTGGT 3038
Qy 2341 GCAGATCCGGACCTGTGAGTATGCGGAGGCGCGACACATATCCCTGGACATTTTC 2400
Db 3039 GCAGATCCGGACCTGTGAGTATGCGGAGGCGCGACACATATCCCTGGACATTTTC 3098
Qy 2401 CGAGGTGGACCCCGCTTACCTCAACCTCTCAGACCTCTGACATATCAAGTACCTCCCAT 2460
Db 3099 CGAGGTGGACCCCGCTTACCTCAACCTCTCAGACCTCTGACATATCAAGTACCTCCCAT 3158
Qy 2461 CGAGTTTATGATCTTAGGAAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 2520
Db 3159 CGAGTTTATGATCTTAGGAAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 3218
Qy 2521 GGCTGAGGAGAGCTGGCCGAGTTCCCGAGCCACAGTGGCCCTGGCCAGGTGAATGGG 2580
Db 3219 GGCTGAGGAGAGCTGGCCGAGTTCCCGAGCCACAGTGGCCCTGGCCAGGTGAATGGG 3278
Qy 2581 CCCCCACGAGGCTGGAGATCACAGAGGAGTCAAGGATGTGGACCGCTGTGGCAGA 2640
Db 3279 CCCCCACGAGGCTGGAGATCACAGAGGAGTCAAGGATGTGGACCGCTGTGGCAGA 3338
Qy 2641 GGCTGCCGTGGGAGGAGGCGCAAGTGGTCTTGGCCGTACGAGAGCTCTTCCACTTCCC 2700
Db 3339 GGCTGCCGTGGGAGGAGGCGCAAGTGGTCTTGGCCGTACGAGAGCTCTTCCACTTCCC 3398
Qy 2701 TGGGAGCACCTGCGCTGGATAGCTGCAGAGCTGGGCTGGCTGAGAGGTGAAGGC 2760
Db 3399 TGGGAGCACCTGCGCTGGATAGCTGCAGAGCTGGGCTGGCTGAGAGGTGAAGGC 3458
Qy 2761 CTCGCTGGAGCACATCTCCGGATCTCTGAAGGGCAGGCCGGAAGGTCTGGAGAGGAGG 2820
Db 3459 CTCGCTGGAGCACATCTCCGGATCTCTGAAGGGCAGGCCGGAAGGTCTGGAGAGGAGG 3518
Qy 2821 GCCCCCAGAGAGAGCAGGCTTGTCTTCTCCGCTCTCAGGTCTGAGAGCTGGGA 2880
Db 3519 GCCCCCAGAGAGAGCAGGCTTGTCTTCTCCGCTCTCAGGTCTGAGAGCTGGGA 3578
Qy 2881 CCGAGGCCGACATTTCTTAAGGAGCTCTCAGATGAGACTGTGGTCTTGGGCCAGTCAGT 2940
Db 3579 CCGAGGCCGACATTTCTTAAGGAGCTCTCAGATGAGACTGTGGTCTTGGGCCAGTCAGT 3638
Qy 2941 GACATGGCTTGCAGGTGTACGCCAGCCAGCTGCCAGGCCACCTGGAGCAAGACGG 3000
Db 3639 GACATGGCTTGCAGGTGTACGCCAGCCAGCTGCCAGGCCACCTGGAGCAAGACGG 3698
Qy 3001 A 3001
Db 3699 A 3699

RESULT 4
US-10-697-263-1
; Sequence 1, Application US/10697263
; Patent No. 6812014
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV2

; CURRENT APPLICATION NUMBER: US/10/697,263
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-697-263-1

Query Match 99.9%; Score 2997.8; DB 4; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCTACCGAGCAGAACTCGGGCCCGAGGCATACAGGGAGCGAGACATCTCTGGCCGCGTGGAG 60
Db 699 CCTACCGAGCAGAACTCGGGCCCGAGGCATACAGGGAGCGAGACATCTCTGGCCGCGTGGAG 758
Qy 61 CCACCCGCTGGTCAACGGGCTGTGACCAAGTTTGAGACCCGCAAGACCCCTCATCTCAT 120
Db 759 CCACCCGCTGGTCAACGGGCTGTGACCAAGTTTGAGACCCGCAAGACCCCTCATCTCAT 818
Qy 121 CTGGAGCTGTGCTCATCCGAGGAGCTGTGGACCCCTGTACAGAAAGGCGCTGGTGAC 180
Db 819 CTGGAGCTGTGCTCATCCGAGGAGCTGTGGACCCCTGTACAGAAAGGCGCTGGTGAC 878
Qy 181 GGAGGCCGAGGTCAGAGGCTTACATCCAGCAGCTGTGGAGGGGCTGCACCTACCTGCACAG 240
Db 879 GGAGGCCGAGGTCAGAGGCTTACATCCAGCAGCTGTGGAGGGGCTGCACCTACCTGCACAG 938
Qy 241 CCATGGCTTCTCACTGGACATAAAGCCCTTAAATCTCTGATGGTGGATCCTGCCCCG 300
Db 939 CCATGGCTTCTCACTGGACATAAAGCCCTTAAATCTCTGATGGTGGATCCTGCCCCG 998
Qy 301 GGAGACATTTAAATCTGGACCTTTGGCTTTGCCAGAGACATCACCCAGCAGAGCTGCA 360
Db 999 GGAGACATTTAAATCTGGACCTTTGGCTTTGCCAGAGACATCACCCAGCAGAGCTGCA 1058
Qy 361 GTTCAGCAGTACGGCTCCCTGAGTTCTCTCCCCGAGATCATCCAGCAGAACCTGT 420
Db 1059 GTTCAGCAGTACGGCTCCCTGAGTTCTCTCCCCGAGATCATCCAGCAGAACCTGT 1118
Qy 421 GAGCAGAGCTCCGACATTTGGGCCATGGGTGTCTCTCTACCTCAGCCTGACCTGCTC 480
Db 1119 GAGCAGAGCTCCGACATTTGGGCCATGGGTGTCTCTCTACCTCAGCCTGACCTGCTC 1178
Qy 481 ATCCCATTTGCCGGGAGAGTGACCGTGCACCCCTCTGAAAGCTCTCTGGAGGGGCGCT 540
Db 1179 ATCCCATTTGCCGGGAGAGTGACCGTGCACCCCTCTGAAAGCTCTCTGGAGGGGCGCT 1238
Qy 541 GTCATGGAGCAGCCCATGGCTGCCACCTCAGCAGAGACGCCCAAAGACTTCAATCAAGGC 600
Db 1239 GTCATGGAGCAGCCCATGGCTGCCACCTCAGCAGAGACGCCCAAAGACTTCAATCAAGGC 1298
Qy 601 TAGCTGCGAGAGAGCCCTCAGGCCCGGCTAGTGGGGCCAGTCTCTCCACCCCTG 660
Db 1299 TAGCTGCGAGAGAGCCCTCAGGCCCGGCTAGTGGGGCCAGTCTCTCCACCCCTG 1358
Qy 661 GTTCTCTGAATCCATGCTCGGAGGAGGCCCATCTTATCAACACCAAGCAGCTCAAGTT 720
Db 1359 GTTCTCTGAATCCATGCTCGGAGGAGGCCCATCTTATCAACACCAAGCAGCTCAAGTT 1418
Qy 721 CCTCTGGCCCGAAGTCGCTGGCAGCGTTTCCCTGATGAGCTACAAGTCAATCTCTGGTAT 780
Db 1419 CCTCTGGCCCGAAGTCGCTGGCAGCGTTTCCCTGATGAGCTACAAGTCAATCTCTGGTAT 1478
Qy 781 GCGCTCATCCCTGAGTGTCTGGGGGCCACCCGACAGCCCTCTCTCGGCTAGGCCG 840

|||||
1479 GCGCTCCATCCCTGAGCTGCTGGGGGCCACCCGACAGCCCTCCCTCGGGGTAGCCCG 1538
QY 841 GCACCTCTGCAAGGACACATGCTGGCTCTCTCAGATTCCTCTCTCTCTGACAAAGAGCT 900
Db 1539 GCACCTCTGCAAGGACACATGCTGGCTCTCTCAGATTCCTCTCTCTCTGACAAAGAGCT 1598
QY 901 CGCCCCATTTCCCGGGCTAAGTCACTGCGCACCTCCCGGGTGACACTCACCACTGCT 960
Db 1599 CGCCCCATTTCCCGGGCTAAGTCACTGCGCACCTCCCGGGTGACACTCACCACTGCT 1658
QY 961 GCACCCCGGGCTTTCTGCGGCCCTCGCGCAGCTGCTGAGGAAGCGGAGGCCAGTGA 1020
Db 1659 GCACCCCGGGCTTTCTGCGGCCCTCGCGCAGCTGCTGAGGAAGCGGAGGCCAGTGA 1718
QY 1021 CGCTTCACAGAGGCCAGCTCGCTCGCTGCACTCTCCGAGGGTCCGGGCCACCGGCCG 1080
Db 1719 CGCTTCACAGAGGCCAGCTCGCTCGCTGCACTCTCCGAGGGTCCGGGCCACCGGCCG 1778
QY 1081 CAGGGCTGCTGCCCGGCCACAGCTCATCCGACAGCTGTTTACACACAGGGCGGGTGA 1140
Db 1779 CAGGGCTGCTGCCCGGCCACAGCTCATCCGACAGCTGTTTACACACAGGGCGGGTGA 1838
QY 1141 GAGCCCTGAGCACCGGGGCTTGCGGCCCGGGGAGCAGCGGCAACCGGCCCGCGGCA 1200
Db 1839 GAGCCCTGAGCACCGGGGCTTGCGGCCCGGGGAGCAGCGGCAACCGGCCCGCGGCA 1898
QY 1201 CTTGCTGAAGGGCGGGTACATTTGGGGGGCGCTGCGAGCTGCGCGAGCCACTGATGA 1260
Db 1899 CTTGCTGAAGGGCGGGTACATTTGGGGGGCGCTGCGAGCTGCGCGAGCCACTGATGA 1958
QY 1261 GCACGGCTGTGGAGGAGGGCGCCAGGAGGAGCAGGCCACCTCTTGGCGCAAGC 1320
Db 1959 GCACGGCTGTGGAGGAGGGCGCCAGGAGGAGCAGGCCACCTCTTGGCGCAAGC 2018
QY 1321 CCCCTCATTTGAGACTGCGCTCCGGCTGCGCTGCGCTGCGCACCCACTTGGGCCCTGGCCA 1380
Db 2019 CCCCTCATTTGAGACTGCGCTCCGGCTGCGCTGCGCTGCGCACCCACTTGGGCCCTGGCCA 2078
QY 1381 CAGCCACTCTCTGAAACATGACTCTCGAGACACCCCGCCCTCTCTCGAGGGCTGCGG 1440
Db 2079 CAGCCACTCTCTGAAACATGACTCTCGAGACACCCCGCCCTCTCTCGAGGGCTGCGG 2138
QY 1441 TGAGGCACAGCACTGCTTACGCCCCCTCGGGGGGGCCCTATCAGGACATGGGGCA 1500
Db 2139 TGAGGCACAGCACTGCTTACGCCCCCTCGGGGGGGCCCTATCAGGACATGGGGCA 2198
QY 1501 CCCTCAGGGCTCCAGCAGCTTCCATCCACTGGTGGCCACCCAGGCACTGCTCAGGCCAGA 1560
Db 2199 CCCTCAGGGCTCCAGCAGCTTCCATCCACTGGTGGCCACCCAGGCACTGCTCAGGCCAGA 2258
QY 1561 GAGGCCATTCCTCCGACAGCCCTTGGGGGACCCAGCCCTTTTCTGCCACCCCAAGCAGGG 1620
Db 2259 GAGGCCATTCCTCCGACAGCCCTTGGGGGACCCAGCCCTTTTCTGCCACCCCAAGCAGGG 2318
QY 1621 TTCTGCCCCCAGAGGGCTCAGCCCCCAGCCAGCAGTGGCCGATGCCCTCTTGGCTC 1680
Db 2319 TTCTGCCCCCAGAGGGCTCAGCCCCCAGCCAGCAGTGGCCGATGCCCTCTTGGCTC 2378
QY 1681 CTTCCCTCAGGATCTTGAAGAGGGCCCTTAGTACCTCAAGCCCTTTTGGGACA 1740
Db 2379 CTTCCCTCAGGATCTTGAAGAGGGCCCTTAGTACCTCAAGCCCTTTTGGGACA 2438
QY 1741 GCCCCAGGACCCCTCTGCTGCAAGAGCCGCCCAATTTGAGCTCTAAGATGGGGCC 1800
Db 2439 GCCCCAGGACCCCTCTGCTGCAAGAGCCGCCCAATTTGAGCTCTAAGATGGGGCC 2498
QY 1801 TGGAGACATCTCTTCTTGGAGGGCCAAAACCGGGCCCTTGAGTTCCCGAGGGTCAGC 1860
Db 2499 TGGAGACATCTCTTCTTGGAGGGCCAAAACCGGGCCCTTGAGTTCCCGAGGGTCAGC 2558
QY 1861 CTCCAGCGGAGCTCTTCCCAAGTGAGCTCCCTCAGGGTGGGCTCTCCCGAGGTGGGCAC 1920
|||||

Db 2559 CTCCACGGCAGACTCTTCCCAAGTGAGCTCCCTCAGGGTGGGCTCCTCCCGAGGTGGGCAC 2618
QY 1921 AGAGCTTGGCCCCCTCCCTCTGATGCGAGGGCTGAGCCCGAGAGGCTGAGGATCTGTCCGA 1980
Db 2619 AGAGCTTGGCCCCCTCCCTCTGATGCGAGGGCTGAGCCCGAGAGGCTGAGGATCTGTCCGA 2678
QY 1981 CTCCACACCCACCTTTCAGAGGGCTCAGGACAGGCGACCATGCGCAAGTTCTCCCTGGG 2040
Db 2679 CTCCACACCCACCTTTCAGAGGGCTCAGGACAGGCGACCATGCGCAAGTTCTCCCTGGG 2738
QY 2041 TGGTTCGGGGGGCTTACGACAGCGTGGCTTGGCTATGCGACCTTTTCCCTTTTGGTGGAGATGC 2100
Db 2739 TGGTTCGGGGGGCTTACGACAGCGTGGCTTGGCTATGCGACCTTTTCCCTTTTGGTGGAGATGC 2798
QY 2101 AGGGGGCATGCTGGGGCAGGGGCCCATGTTGGCCAGGATAGCTTGGGCTGTGTCCAGTC 2160
Db 2799 AGGGGGCATGCTGGGGCAGGGGCCCATGTTGGCCAGGATAGCTTGGGCTGTGTCCAGTC 2858
QY 2161 GGAGGAGGAGCAGGAGGAGGAGGCGCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2220
Db 2859 GGAGGAGGAGCAGGAGGAGGAGGAGGCGCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2918
QY 2221 CAGGGCTGAGAGGCCACTTGGCCCCCAGGTCAGTGCAGAGGCTGTGTCTGAGGTTCGGCAGGGC 2280
Db 2919 CAGGGCTGAGAGGCCACTTGGCCCCCAGGTCAGTGCAGAGGCTGTGTCTGAGGTTCGGCAGGGC 2978
QY 2281 TCCACACAGAGGCTCTCCAGAGCCCAACCCATGAGGAGACATCGGGCAGGTCTCTCCCTGGT 2340
Db 2979 TCCACACAGAGGCTCTCCAGAGCCCAACCCATGAGGAGACATCGGGCAGGTCTCTCCCTGGT 3038
QY 2341 GCAGATCCGGGACCTGTGAGTGTGCGAGGGCGGCGACACATATATCCCTTGGACATTTTC 2400
Db 3039 GCAGATCCGGGACCTGTGAGTGTGCGAGGGCGGCGACACATATATCCCTTGGACATTTTC 3098
QY 2401 CGAGGTGGAACCCCGCTTACCTCAACCTCTCAGACCTGTAGGATATCAAGTACCTCCCATTT 2460
Db 3099 CGAGGTGGAACCCCGCTTACCTCAACCTCTCAGACCTGTAGGATATCAAGTACCTCCCATTT 3158
QY 2461 CGAGTTTATGATCTTTCAGGAAAGTCCCAAGTCCGCTCAGCAGAGCGGCCCTCCGCCAT 2520
Db 3159 CGAGTTTATGATCTTTCAGGAAAGTCCCAAGTCCGCTCAGCAGAGCGGCCCTCCGCCAT 3218
QY 2521 GGCTGAGGAGGAGTGGCGCGAGTTCCCGGAGCCCAAGTGGCCCTTGGCCAGGTGAACCTGGG 2580
Db 3219 GGCTGAGGAGGAGTGGCGCGAGTTCCCGGAGCCCAAGTGGCCCTTGGCCAGGTGAACCTGGG 3278
QY 2581 CCCCCAGCAGGCTGTGAGATCACAGAGGAGTCAAGAGATGTGAGCGCTGTCTGGGAGA 2640
Db 3279 CCCCCAGCAGGCTGTGAGATCACAGAGGAGTCAAGAGATGTGAGCGCTGTCTGGGAGA 3338
QY 2641 GGCTGCGTGGGACGAGGAGCGCAAGTGGTCTTCCGCTCAGCAGGCTCTTCCACTTTCCC 2700
Db 3339 GGCTGCGTGGGACGAGGAGCGCAAGTGGTCTTCCGCTCAGCAGGCTCTTCCACTTTCCC 3398
QY 2701 TGGGAGGACCTGCGCTGGATGAGCTGAGAGCTGGGGCTGGCTGAGAGGTGAAGGC 2760
Db 3399 TGGGAGGACCTGCGCTGGATGAGCTGAGAGCTGGGGCTGGCTGAGAGGTGAAGGC 3458
QY 2761 CTCCGTGGAGCAGATCTCCCGGATCTCTGAAGGGCAGGCCCGGAGGTCTGGAAGAGGAGGG 2820
Db 3459 CTCCGTGGAGCAGATCTCCCGGATCTCTGAAGGGCAGGCCCGGAGGTCTGGAAGAGGAGGG 3518
QY 2821 GCCCCCAGGAAAGCAGGCCCTTGTCTCTTCCGGCTCTCAGGTCTGAAGAGCTGGGA 2880
Db 3519 GCCCCCAGGAAAGCAGGCCCTTGTCTCTTCCGGCTCTCAGGTCTGAAGAGCTGGGA 3578
QY 2881 CCGAGGCCACATCTCTTAAGGAGGCTCTCAGATGAGATGTGTCTTGGGCCAGTCACT 2940
Db 3579 CCGAGGCCACATCTCTTAAGGAGGCTCTCAGATGAGATGTGTCTTGGGCCAGTCACT 3638
QY 2941 GACACTGGCTTCCAGGTGTACGCCAGCAGCTTCCCGAGGCCACCTTGGGACAAAGACGG 3000
Db 3639 GACACTGGCTTCCAGGTGTACGCCAGCAGCTTCCCGAGGCCACCTTGGGACAAAGACGG 3698
|||||

Qy 3001 A 3001
Db 3699 A 3699

RESULT 5
US-10-697-263-3
; Sequence 3, Application US/10697263
; Patent No. 6812014
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV2
; CURRENT APPLICATION NUMBER: US/10/697,263
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-697-263-3

Query Match 99.9%; Score 2997.8; DB 4; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCTACGGAGCAGAACTCGGGCCCAAGGATACAGGGAGCGAGACATCTCTGGCCGGCTGAG 60
Db 699 CCTACGGAGCAGAACTCGGGCCCAAGGATACAGGGAGCGAGACATCTCTGGCCGGCTGAG 758

Qy 61 CCACCCGCTGTACAGGGGCTGTGGACCACTTTGAGACCGCAAGACCTCATCTCTCAT 120
Db 759 CCACCCGCTGTACAGGGGCTGTGGACCACTTTGAGACCGCAAGACCTCATCTCTCAT 818

Qy 121 CCTGGAGCTGTCTATCCGAGGAGCTGTGGACCGCTGTACAGGAAGGCGGTGTGAC 180
Db 819 CCTGGAGCTGTCTATCCGAGGAGCTGTGGACCGCTGTACAGGAAGGCGGTGTGAC 878

Qy 181 GGAGGCGAGGTGAAGGTCTACATCCAGAGCTGTGGAGGGGCTGCATCTACCTGCACAG 240
Db 879 GGAGGCGAGGTGAAGGTCTACATCCAGAGCTGTGGAGGGGCTGCATCTACCTGCACAG 938

Qy 241 CCATGGGGTTCTCCACCTGGACATAAAGCCCTCTAAACATCTGTATGTGTGATCTGCCCCG 300
Db 939 CCATGGGGTTCTCCACCTGGACATAAAGCCCTCTAAACATCTGTATGTGTGATCTGCCCCG 998

Qy 301 GGAAGACATTAAATCTGCGACTTTGGCTTTGCCAGAAACATACACCCAGCAGAGCTGCA 360
Db 999 GGAAGACATTAAATCTGCGACTTTGGCTTTGGCTTTGCCAGAAACATACACCCAGCAGAGCTGCA 1058

Qy 361 GTTCAGGACGTACGGCTCCCTGTAGTTGCTTCCCGGAGATCATTCAGCAGAAACCTGT 420
Db 1059 GTTCAGGACGTACGGCTCCCTGTAGTTGCTTCCCGGAGATCATTCAGCAGAAACCTGT 1118

Qy 421 GAGCGAAGCTCCGACATTTGGGCGCATGGGTGTCTATCTCTACCTAGCGCTGACCTGCTC 480
Db 1119 GAGCGAAGCTCCGACATTTGGGCGCATGGGTGTCTATCTCTACCTAGCGCTGACCTGCTC 1178

Qy 481 ATCCCCATTTCCGGCGAGAGTGACCGGTGCCACCTCTCTGAACGTCTCTGAGGGGGCGGT 540
Db 1179 ATCCCCATTTCCGGCGAGAGTGACCGGTGCCACCTCTCTGAACGTCTCTGAGGGGGCGGT 1238

Qy 541 GTCATGAGCAGGCCCCATGGCTGCCCACTTCAGCGAAGACGCCAAGACTTCATCAAGGC 600

Db 1239 GTCATGAGCAGGCCCCATGGCTGCCACCTCAGCGAAGACGCCAAAGACTTCATCAAGGC 1298
Qy 601 TAGCTGTCAGAGAGCCCTCAGGCCGGCCCTAGTGGGGCCAGTGCTCTCTCCACCCCTG 660
Db 1299 TAGCTGTCAGAGAGCCCTCAGGCCGGCCCTAGTGGGGCCAGTGCTCTCTCCACCCCTG 1358
Qy 661 GTTCTGAAATCCATGCTCGGAGGAGGCCACTTCATCAACACCAAGCAGCTCAAGTT 720
Db 1359 GTTCTGAAATCCATGCTCGGAGGAGGCCACTTCATCAACACCAAGCAGCTCAAGTT 1418
Qy 721 CTTCTGGGGCCGAAGTCCCTGGCAGCGTTCCCTGATGAGCTACAAGTCAATCTCTGTGAT 780
Db 1419 CTTCTGGGGCCGAAGTCCCTGGCAGCGTTCCCTGATGAGCTACAAGTCAATCTCTGTGAT 1478
Qy 781 GCGCTCCATCCCTGAGCTGCTGGGGCCCAACCCGACAGCCCTCCCTCGGGCTAGGCCG 840
Db 1479 GCGCTCCATCCCTGAGCTGCTGGGGCCCAACCCGACAGCCCTCCCTCGGGCTAGGCCG 1538
Qy 841 GCACCTCTGACAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTGACAAACGAGCT 900
Db 1539 GCACCTCTGACAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTGACAAACGAGCT 1598
Qy 901 CGCCCCATTTGCCGGGCTAAGTCACTGCCACCTCCCGGGTGACACACTCACCACCTGCT 960
Db 1599 CGCCCCATTTGCCGGGCTAAGTCACTGCCACCTCCCGGGTGACACACTCACCACCTGCT 1658
Qy 961 GCACCCCGGGGCTTCTCGGGCCCTCGGGCAGCTGCTGAGGAAGCCGAGGCGAGTGA 1020
Db 1659 GCACCCCGGGGCTTCTCGGGCCCTCGGGCAGCTGCTGAGGAAGCCGAGGCGAGTGA 1718
Qy 1021 GCCTCTCACCGAGGCCCACTCCGCTGCTATCTCCGAGGGTGCCTGGGCGCACCGGCCG 1080
Db 1719 GCCTCTCACCGAGGCCCACTCCGCTGCTATCTCCGAGGGTGCCTGGGCGCACCGGCCG 1778
Qy 1081 CAGGGCTGGTCCCGGGCAGAGCTCATCCGAGGCTGTTTACACACAGGGGGTGA 1140
Db 1779 CAGGGCTGGTCCCGGGCAGAGCTCATCCGAGGCTGTTTACACACAGGGGGTGA 1838
Qy 1141 GAGCCCTGAGCAGCGGGCCCTGGCCCGGGGAGCAGCGGCACCCCGCCCGGGGGCGCA 1200
Db 1839 GAGCCCTGAGCAGCGGGCCCTGGCCCGGGGAGCAGCGGCACCCCGCCCGGGGGCGCA 1898
Qy 1201 CTTGCTGAAGGGCGGGTACATTCGGGGGGCGCTGTCAGGGCTCGCGAGCACTGATGA 1260
Db 1899 CTTGCTGAAGGGCGGGTACATTCGGGGGGCGCTGTCAGGGCTCGCGAGCACTGATGA 1958
Qy 1261 GCACCGGTGCTGAGGAGGAGCGCCGAGGAGGAGCAGGCCACCTCTCTGCGCAAGC 1320
Db 1959 GCACCGGTGCTGAGGAGGAGCGCCGAGGAGGAGCAGGCCACCTCTCTGCGCAAGC 2018
Qy 1321 CCCTCATTCGAGACTGCCCTCCGGCTGCTGCTCTGGCACCCACCTTTGGCCCCCTGGCCA 1380
Db 2019 CCCTCATTCGAGACTGCCCTCCGGCTGCTGCTCTGGCACCCACCTTTGGCCCCCTGGCCA 2078
Qy 1381 CAGCCACTCCCTGGAACATGACTCTCGAGCAGCCCCCGCCCTCTCTCGAGGCGCTCGG 1440
Db 2079 CAGCCACTCCCTGGAACATGACTCTCGAGCAGCCCCCGCCCTCTCTCGAGGCGCTCGG 2138
Qy 1441 TGAGGCACAGCAGACTGCTTACGCCCTCCGGGGGGGGCCCTATCAGGACATGGGGCA 1500
Db 2139 TGAGGCACAGCAGACTGCTTACGCCCTCCGGGGGGGGCCCTATCAGGACATGGGGCA 2198
Qy 1501 CCCTCAGGGTCCCAAGCAGCTTCCATCCACTGTGTGGCCACCCAGGCACTGCTCAGCCAGA 1560
Db 2199 CCCTCAGGGTCCCAAGCAGCTTCCATCCACTGTGTGGCCACCCAGGCACTGCTCAGCCAGA 2258
Qy 1561 GAGGCCATCCCGGAGCAGCCCTTGGGGGGCAGCCAGCCCTTTCTGCCACCCCAAGCAGGG 1620
Db 2259 GAGGCCATCCCGGAGCAGCCCTTGGGGGGCAGCCAGCCCTTTCTGCCACCCCAAGCAGGG 2318
Qy 1621 TTCTGCCCCCAGGAGGGCTGACGCCCCACCCAGCAGTTGCCCCCATGCTCTCTGCTC 1680

Db 2319 TTCTGCCCCCAGGAGGGCTGCAGCCCCCAGCAGAGTTGTCCTCCATGCTCCTGCTC 2378
Qy 1681 CTTTCCTCCAGGATCTTGAAGAGGGCCCCCTTAGTACCCTCAAGAGCCCTTCTTGGGACA 1740
Db 2379 CTTTCCTCCAGGATCTTGAAGAGGGCCCCCTTAGTACCCTCAAGAGCCCTTCTTGGGACA 2438
Qy 1741 GCCCAGGACCCCTGCTCCCTGCAAGAGAGCCGCCCATTTGGACTCTAAGATGGGGCC 1800
Db 2439 GCCCAGGACCCCTGCTCCCTGCAAGAGAGCCGCCCATTTGGACTCTAAGATGGGGCC 2498
Qy 1801 TGGAGACATCTCTCTTCTGGAGGCCCCAAACCCGGGCCCTGCAAGTTCCCGAGGGTCAGC 1860
Db 2499 TGGAGACATCTCTCTTCTGGAGGCCCCAAACCCGGGCCCTGCAAGTTCCCGAGGGTCAGC 2558
Qy 1861 CTCCAGGAGGCTCTTCCCAAGTAGCTCCTCAGGGTGGGCTCTCCCAAGTGGGAC 1920
Db 2559 CTCCAGGAGGCTCTTCCCAAGTAGCTCCTCAGGGTGGGCTCTCCCAAGTGGGAC 2618
Qy 1921 AGAGCCTGGGCCCTCTGATGCGAGGGCTGACCCAGGAGCTGAGGATCTGCCGA 1980
Db 2619 AGAGCCTGGGCCCTCTGATGCGAGGGCTGACCCAGGAGCTGAGGATCTGCCGA 2678
Qy 1981 CTCACACCCACCTTGAGCGGCCCTCAGGAACAGGCGACCATGCGCAAGTTCTCCCTGGG 2040
Db 2679 CTCACACCCACCTTGAGCGGCCCTCAGGAACAGGCTGACCATGCGCAAGTTCTCCCTGGG 2738
Qy 2041 TGGTCGGGGGGCTACGAGGCGTGGCTGGCTATGCGACCTTTGCTGGTGGAGATGC 2100
Db 2739 TGGTCGGGGGGCTACGAGGCGTGGCTGGCTATGCGACCTTTGCTGGTGGAGATGC 2798
Qy 2101 AGGGGGATGCTGGGGCAGGGGCCCATGTGGGGCAGGATAGCTGGGCTGTGCCAGTC 2160
Db 2799 AGGGGGATGCTGGGGCAGGGGCCCATGTGGGGCAGGATAGCTGGGCTGTGCCAGTC 2858
Qy 2161 GGAGGAGGAGCAGGAGGCGGCGAGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 2220
Db 2859 GGAGGAGGAGCAGGAGGCGGCGAGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 2918
Qy 2221 CAGGGCTGAGAGCCACTGCCCCAGGTCAGTGCAGAGGCTGTGCTGAGGTCGCGAGGC 2280
Db 2919 CAGGGCTGAGAGCCACTGCCCCAGGTCAGTGCAGAGGCTGTGCTGAGGTCGCGAGGC 2978
Qy 2281 TCCACAGGAGCTCTCCAGAGCCACCCCATGGAGGACATCGGGAGGCTCTCCCTGGT 2340
Db 2979 TCCACAGGAGCTCTCCAGAGCCACCCCATGGAGGACATCGGGAGGCTCTCCCTGGT 3038
Qy 2341 GCAGATCGGGACCTGTGAGGTGATGGGAGGCGCGACACATATCCCTGGACATTTTC 2400
Db 3039 GCAGATCGGGACCTGTGAGGTGATGGGAGGCGCGCGACACATATCCCTGGACATTTTC 3098
Qy 2401 CGAGGTGGAACCCGCCCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db 3099 CGAGGTGGAACCCGCCCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 3158
Qy 2461 CGAGTTTATGATCTTCAGGAAGTCCCAAGTCCGCTCAGCAGAGCGGCCCTCCCCCAT 2520
Db 3159 CGAGTTTATGATCTTCAGGAAGTCCCAAGTCCGCTCAGCAGAGCGGCCCTCCCCCAT 3218
Qy 2521 GGCTGAGGAGGAGTGGCCGAGTTCCCGAGGCCACAGTGGCCCTGGCCAGGTGAACCTGGG 2580
Db 3219 GGCTGAGGAGGAGTGGCCGAGTTCCCGAGGCCACAGTGGCCCTGGCCAGGTGAACCTGGG 3278
Qy 2581 CCCCCAGCGGCTGGAGATCACAGAGGAGTCAGAGGATGTGGACGCGCTGTCTGGCAGA 2640
Db 3279 CCCCCAGCGGCTGGAGATCACAGAGGAGTCAGAGGATGTGGACGCGCTGTCTGGCAGA 3338
Qy 2641 GGCTGCCGTGGGAGGAGCGCAAGTGGTCTCGCGCTCAGCGAGCTCTTCCACTTCCC 2700
Db 3339 GGCTGCCGTGGGAGGAGCGCAAGTGGTCTCGCGCTCAGCGAGCTCTTCCACTTCCC 3398
Qy 2701 TGGGAGGACCTCGCGCTGGATGAGCTTGAGAGCTGGGGCTGGCTGAGAGGTGAAGGC 2760
Db 3399 TGGGAGGACCTCGCGCTGGATGAGCTTGAGAGCTGGGGCTGGCTGAGAGGTGAAGGC 3458

Qy 2761 CTCCTGGAGCAGCATCTCCGGATCTTGAAGGCGAGCCCGAGGCTCTGGAGAGGAGGG 2820
Db 3459 CTCCTGGAGCAGCATCTCCGGATCTTGAAGGCGAGCCCGAGGCTCTGGAGAGGAGGG 3518
Qy 2821 GCCCCCCAGGAAGAGCCAGGCTTGTCTTCTCCGGCTCTCAGGCTCTGAAGAGCTGGGA 2880
Db 3519 GCCCCCCAGGAAGAGCCAGGCTTGTCTTCTCCGGCTCTCAGGCTCTGAAGAGCTGGGA 3578
Qy 2881 CGAGGCCCGACATCTCTAAGGAGGCTCTCAGATGAGACTGTGTCTCTGGGCCAGTCACT 2940
Db 3579 CGAGGCCCGACATCTCTAAGGAGGCTCTCAGATGAGACTGTGTCTCTGGGCCAGTCACT 3638
Qy 2941 GACACTGGCCTGCGAGGTGTGAGCCAGGCTGCGCCAGGACCTGGAGCAAGAGCGG 3000
Db 3639 GACACTGGCCTGCGAGGTGTGAGCCAGGCTGCGCCAGGACCTGGAGCAAGAGCGG 3698
Qy 3001 A 3001
Db 3699 A 3699

RESULT 6

US-09-159-385-4
; Sequence 4, Application US/09159385
; Patent No. 5958748
; GENERAL INFORMATION:
; APPLICANT: AKIRA, SHIZUO
; APPLICANT: KAWAI, TARO
; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
; FILE REFERENCE: PH-569
; CURRENT APPLICATION NUMBER: US/09/159,385
; EARLIER FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: JP97/261589
; EARLIER FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1429
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)..(1353)
US-09-159-385-4

Query Match 5.7%; Score 172; DB 2; Length 1429;
Best Local Similarity 54.4%; Pred. No. 5.1e-27;
Matches 371; Conservative 0; Mismatches 305; Indels 6; Gaps 1;

Qy 17 CGGGCCCGAGCATACAGGGAGCGAGACATCTTGCCCGCGCTGAGCCACCCGCTGGTCAG 76
Db 181 CGGGAGGAGATCGAAACGCGAGGTGAGCATCTGCGCGAGATCCGCCACCCCAACATCATA 240
Qy 77 GGGCTGCTGAGACAGTTTGAAGCCCGAGACCTCATCTCATCTCTGGAGCTGTGTCA 136
Db 241 ACATGTCATGACGTGTTTCGAGAAACAGACAGATGTGTGTGTGATCTCTGGAGCTGGTGTCC 300
Qy 137 TCCGAGGAGCTGTGGACCCCTGTGTACAGGAAGGGCGTGTGACGGAGGCGAGGCTCAAG 196
Db 301 GGTGGCGAGCTTTCGACTTCTTGGCCGAGAGAGGATCATTTGACGGAGGATGAGGCCACG 360
Qy 197 GTCTACATCCAGAGCTGTGGAGGGCTGCACTTACTGTCACAGCCATGGCGTTCTCCAC 256
Db 361 CAGTTCTCTAAACAAATCTTAGACGGTGTCCACTTACCTGACATCTCAAGCGCATCGCACAC 420
Qy 257 CTGGACATAAAGCCCTTAACATCTGATGGT-----GCATCTCTCCCGGGAAGACATT 310
Db 421 TTTGACCTGAAGCCCGGAGAACATCATGTTGTCGACAAAGCAGCAGCGCCAGCCCGGCATT 480
Qy 311 AAATCTGCGACTTTGGCTTTGCCAGAACATCACCCCGAGCAGAGCTGCAGTTTCAGCCAG 370
Db 481 AAGCTCATGCACTTTGGCATCGCGCACAGGATCGAGCTGGCAGCGAGTTCAAGACATC 540

Qy	371	TACGGCTCCCTGATGTTTCGTCTCCCCCGAGATCATCCAGCAGAAACCTGTGTAGCGAAGCC	430
Db	541	TTTGGCACACCCGAGTTTGTTCGCCCCGAGATCGTGAACATATGAGCACTTGGCTTGGAG	600
Qy	431	TCCGACATTTTGGGCCATGGGTGTCTATCTCTTACCTCAGCTGCACTGCTCATCCCATTT	490
Db	601	GCTGACATGTGGAGCATTTGGGGTTCATCACCTACATCTCTGAGGGGAGCGTCCCATTC	660
Qy	491	GCCGGCGAGAGTGCAGTGCACCGTCCACCTCTCTGAACGTCTTGGAGGGGGCGGTGCATGGAGC	550
Db	661	CTGGCGGAGACCAAGCAGGAGACCGTGCAGAACATCTCAGCAGTGAATATGACTTTGAT	720
Qy	551	AGCCCCATGGTGGCTGCCACTCAGCGAAGACGCCAAAGACTTTCATCAAGCTACGCTGCAG	610
Db	721	GAGGAATACTTCAGGAGACACAGCGAGCTGGCCAAAGGACTTCATCCGAGGCTGCTGTGTC	780
Qy	611	AGAGCCCTTCAGGCCCGGCCCTAGTTCGGGCCCAAGTGCCTCTCCACCCCTGGTTCCTGAAA	670
Db	781	AAAGACCCCAAGAGAGGATGACCATGCACAGAGCCTGGAGCAITCTCTGGATCAAGGTG	840
Qy	671	TCCATGCTGCGGAGGAGGCC	692
Db	841	CGCAGGGCGGAGGACGGCGCCC	862

```

RESULT 7
US-09-186-277-4
; Sequence 4, Application US/09186277
; Patent No. 6171841
; GENERAL INFORMATION:
; APPLICANT: KAWAI, TARO
; APPLICANT: AKIRA, SHIZUO
; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
; FILE REFERENCE: 081356/0128
; CURRENT APPLICATION NUMBER: US/09/186,277
; CURRENT FILING DATE: 1998-11-05
; EARLIER APPLICATION NUMBER: JP97/261589
; EARLIER FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1429
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10) .. (1353)
; US-09-186-277-4

```

QY	311	AAATCTCGGACTTTGGCTTTGCCAGAAATCATCAACCCAGCAGAGCTCGAGTTCAAGCCAG	370
DB	481	AAGCTCATCGACTTTGGCATCGCGACAGGATCGAGGCTGGCAGCGAGTTCAAGAACATC	540
QY	371	TACGGCTCCCTGAGTTGCTCTCCCGAGATCATCCAGCAGAACCCCTGTGAGCGAAGCC	430
DB	541	TTTGCGCACACCCGAGTTGTTCGCCCCGAGATCGTGAACATATGAGCCACTTGGCTTGGAG	600
QY	431	TCCGACATTTGGGCGCATGGGTGTCTATCTCTTACCTTCAGCCCTGACCTGCTCATCCCCATTT	490
DB	601	GCTGACATGTGAGCATTTGGCGTTCATCACTTACATCTCTCTGAGGAGCGTCCCCATTC	660
QY	491	GCCGGCGAGAGTAGCCGTGCCACCCCTCTGAACTCTCTGAGGGGGCGGTCTATGAGAGC	550
DB	661	CTGGCGGAGAGCAAGACGAGGACGCTGACGAACATCTCAGCAGTGAACATATGACTTTGAT	720
QY	551	AGCCCCATGGCTGCCCACTCTCAGCGAAGACGCCAAAGACTTTCATCAAGGCTACGCTGCAG	610
DB	721	GAGAAATCTTCAGAGGACACGAGGAGCTGTGCGCAAGACTTTCATCCGAGGCTGCTGTCTC	780
QY	611	AGAGCCCTCTCAGGCCCGGCTTAGTGCGGCCCAAGTGCCTCTCCCAACCCCTGGTTTCTGAAA	670
DB	781	AAAGACCCCAAGAGGAGGATGACCATCGCACAGAGCTTGGAGCATTCCTGGATCAAGGTG	840
QY	671	TCCATGCTCGGAGGAGGCC	692
DB	841	CGCAGGCGGAGGACGGCGCC	862

RESULT 8
 US-09-949-016-1966
 ; Sequence 1966, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CLO01307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1966
 ; LENGTH: 2157
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-1966

```

RESULT 8
US-09-949-016-1966
; Sequence 1966, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1966
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1966

```

Query Match	5.7%	Score 170.2;	DB 4;	Length 2157;
Best Local Similarity	55.0%;	Pred. No. 1.4e-26;		
Matches 359;	Conservative 0;	Mismatches 288;	Indels 6;	Gaps 1;
QY	17	CGGCCCCAGGCATACAGGGACGAGACATCTCTGCCGCCGCTGAGCCACCCGCTGGTCAAG	76	
DB	317	CGGAGGAGATCGAGCGGGAGGTGAACATCTCGGGAGATCCGGCACCCCAACATCATC	376	
QY	77	GGGCTGCTGACACGATTTGAGACCGCGAAGACCTCATCTCATCTGGAGCTGTGTCTCA	136	
DB	377	ACCTTGACGACATCTTCGAGAACAGACGAGCTGTCTCTCATCTCTGGAGCTGGTCTCT	436	
QY	137	TCCGAGGAGCTGTGGACCGCCTGTCACGAAAGCGCGCTGTGTGACGGAGGCCGAGGTCAAG	196	
DB	437	GGCGGGGAGCTCTTTGTACTTCTCGCGGAGAAAGAGTCTGTGACGGAGGACGAGGCCACC	496	
QY	197	GTCTACATCCAGACAGCTGTGTGGGGGCTGCACCTACCTGCACAGCCCATGGGCTTCTCCAC	256	

Db 497 CAGTTCTCAAGCAGATCTCTGACGGGTTTCACTACCTGCACTCTAAGCGCATCGCAC 556
Qy 257 CTGACATATAAGCCCTCTAATCATCTCTGATGCTGATCTCTGCGGGAAGAC-----ATT 310
Db 557 TTTGACCTGAAGCCGGAACATCATCTGCTGACAAAGACGTCGCCAACCCACGAATC 616
Qy 311 AAAATCTGCGACTTTGGCTTTGGCCAGAAATACACCCAGCAGAGCTGCAGTTCCAGCCAG 370
Db 617 AAGCTCATCGACTTCGGCATCGCGCAAGATCAGGGGGGGAACGAGTTCAAGAAATC 676
Qy 371 TAGCGCTCCCTGAGTTTGGTCTCCCGGAGATCATCCAGCAAGAACCCCTGTGAGGGAAGCC 430
Db 677 TTCGGCACCCCGGAGTTTGGCCCGCAGAGATTGTGAATATGAGCCGCTGGGCTGGAG 736
Qy 431 TCCGACATTTGGGCGCATGGGTGTCTCTCTACCTCAGCCTGACCTGCTCATCCCATTTT 490
Db 737 GCGGACATGTGGAGCATCGGTGTCTCATCACTATATCTCTCCTGAGCGGTGTCATCCCGTTC 796
Qy 491 GCCGGCAGAGTACCGTGCACCTCTCTGAACGTCTGGAGGGGCGGTGTCTATGGAGC 550
Db 797 CTGGCGAGACCAAGCAGGAGACGCTTACCAACATCTCAGCCGTGAATCAGACTTCGAC 856
Qy 551 AGCCCCATGGTGGCCACCTCAGCGAAGACGCCAAAGACTTTCATCAAGGCTTACGCTGCAG 610
Db 857 GAGGAGTACTTCAGCAACACACGAGCGAGCTGSCCAAGGACTTTCATTGCGCGGCTGCTGTC 916
Qy 611 AGAGCCCTCAGGCGCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTGTTT 663
Db 917 AAAGATCCCAAGCGGAGAAATGACCATTTGCCAGAGCCTGGAACATTCCTGGAT 969

RESULT 9

US-09-949-016-195
; Sequence 195, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195
; LENGTH: 2105
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195

Query Match 5.6%; Score 168.6; DB 4; Length 2105;
Best Local Similarity 54.8%; Pred. No. 3e-26;
Matches 358; Conservative 0; Mismatches 289; Indels 6; Gaps 1;

Qy 17 CGGCCCCAGGCATACAGGAGCGGAGATCTCTGCGCGGCTGAGCCACCCCTGGTCAAG 76
Db 265 CGGAGGAGATCGAGCGGGAGGTGAACATCTCTGGGAGATCCGGCACCCCAACATCATC 324
Qy 77 GGGCTGTGGACCATTTGAGACCGCAAGACCTCATCTCTCTGAGCTGTGCTCA 136
Db 325 ACCCTGCACGACATCTTCGAGAACAGACGAGCTGTCTCTCTCTGAGCTGTCTCT 384
Qy 137 TCCGAGGAGCTGTGGACCGCTGTACAGGAAGGGCGTGTGACGGAGCGCGAGGTCAAG 196
Db 385 GCGGGGAGCTCTTTGACTTCTCTGCGGAGAAAGAGTCGCTGACGGAGGACGAGGCCACC 444
Qy 197 GTCTACATCCAGCAGCTGGTGGAGGGGCTGCATCTGACAGCCATGGCGGTCTCTCAC 256

Db 445 CAGTTCTCAAGCAGATCTCTGACGGGTTTCACTACCTGCACTCTAAGCGCATCGCAC 504
Qy 257 CTGACATATAAGCCCTCTAATCATCTCTGATGCTGATCTCTGCGGGAAGAC-----ATT 310
Db 505 TTTGACCTGAAGCCGGAACATCATCTGCTGACAAAGACGTCGCCAACCCACGAATC 564
Qy 311 AAAATCTGCGACTTTGGCTTTGGCCAGAAATACACCCAGCAGAGCTGCAGTTCCAGCCAG 370
Db 565 AAGCTCATCGACTTCGGCATCGCGCAAGATCAGGGGGGGAACGAGTTCAAGAAATC 624
Qy 371 TAGCGCTCCCTGAGTTTGGTCTCCCGGAGATCATCCAGCAAGAACCCCTGTGAGGGAAGCC 430
Db 625 TTCGGCACCCCGGAGTTTGGCCCGCAGAGATTGTGAATATGAGCCGCTGGGCTGGAG 684
Qy 431 TCCGACATTTGGGCGCATGGGTGTCTCTCTACCTCAGCCTGACCTGCTCATCCCATTTT 490
Db 685 GCGGACATGTGGAGCATCGGTGTCTCATCACTATATCTCTCCTGAGCGGTGTCATCCCGTTC 744
Qy 491 GCCGGCAGAGTACCGTGCACCTCTCTGAACGTCTGGAGGGGCGGTGTCTATGGAGC 550
Db 745 CTGGCGAGACCAAGCAGGAGACGCTTACCAACATCTCAGCCGTGAATCAGACTTCGAC 804
Qy 551 AGCCCCATGGTGGCCACCTCAGCGAAGACGCCAAAGACTTTCATCAAGGCTTACGCTGCAG 610
Db 805 GAGGAGTACTTCAGCAACACACGAGCGAGCTGSCCAAGGACTTTCATTGCGCGGCTGCTGTC 864
Qy 611 AGAGCCCTCAGGCGCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTGTTT 663
Db 865 AAAGATCCCAAGCGGAGAAATGACCATTTGCCAGAGCCTGGAACATTCCTGGAT 917

RESULT 10

US-09-159-385-3
; Sequence 3, Application US/09159385
; Patent No. 5958748
; GENERAL INFORMATION:
; APPLICANT: AKIRA, SHIZUO
; APPLICANT: KAWAI, TARO
; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
; FILE REFERENCE: PH-569
; CURRENT APPLICATION NUMBER: US/09/159,385
; CURRENT FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: JP97/261589
; EARLIER FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2132
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (94)..(1455)
US-09-159-385-3

Query Match 5.6%; Score 168.6; DB 2; Length 2132;
Best Local Similarity 54.8%; Pred. No. 3e-26;
Matches 358; Conservative 0; Mismatches 289; Indels 6; Gaps 1;

Qy 17 CGGCCCCAGGCATACAGGAGCGGAGATCTCTGCGCGGCTGAGCCACCCCTGGTCAAG 76
Db 265 CGGAGGAGATCGAGCGGGAGGTGAACATCTCTGGGAGATCCGGCACCCCAACATCATC 324
Qy 77 GGGCTGTGGACCATTTGAGACCGCAAGACCTCATCTCTCTGAGCTGTGCTCA 136
Db 325 ACCCTGCACGACATCTTCGAGAACAGACGAGCTGTCTCTCTCTGAGCTGTCTCT 384
Qy 137 TCCGAGGAGCTGTGGACCGCTGTACAGGAAGGGCGTGTGACGGAGCGCGAGGTCAAG 196
Db 385 GCGGGGAGCTCTTTGACTTCTCTGCGGAGAAAGAGTCGCTGACGGAGGACGAGGCCACC 444
Qy 197 GTCTACATCCAGCAGCTGGTGGAGGGGCTGCATCTGACAGCCATGGCGGTCTCTCAC 256

Db 445 CAGTTCCTCAAGCAGATCCTGGACGGGTTCACTACCTGCACTCTAAGCCGATCGCAC 504
Qy 257 CTGACATAAAGCCCTCTAAACATCCTGATGGTGATCCTGCGCGGAAGAC-----ATT 310
Db 505 TTTGACCTGAAGCCGGAACATCATCTGCTGACAAAGACGTGCCAACCCACGAATC 564
Qy 311 AAAATCTGGCAGCTTTGGCTTTGCCAGAAATCATCACCCAGCAGAGCTGCAAGTTCAGCCAG 370
Db 565 AAGCTCATCGACTTCGGCATCGCGACAAGATCGAGGCGGGAACAGATTCAAGAAACATC 624
Qy 371 TAGGGCTCCCTGAGTTGCTCTCCCGAGATCATCCAGAGACCTGTGAGGGAAGCC 430
Db 625 TTCGGCACCCCGGAGTTTGGCCCCAGAGATTGTGAATATGAGCCGCTGGGCGCTGGAG 684
Qy 431 TCCGACATTTGGGCATAGGGTGTCTCTACCTCAGCCTGACCTGCTCATCCCCATT 490
Db 685 GCGGACATGTGGACATCGGTGTCTATACCTATATCTCTGAGCGGTGATCCCCGTTT 744
Qy 491 GCCGGCAGAGTGAACCTCTGAAACGTCTGAGGGGCGCGTGTATGAGGC 550
Db 745 CTGGCGAGACCAAGCAGGAGAGCTCACCAACATCTCAGCCGCTGAACTACGACTTCGAC 804
Qy 551 AGCCCCATGGTCCGACCTCAGCGAAGACGCCAAGACTTTCATCAAGGCTACGCTGCAG 610
Db 805 GAGGAGTACTTCAGCAACACCGAGCGAGCTGGCCAAAGGACTTCAATTCGCCGCGTCTCGTC 864
Qy 611 AGAGCCCTCAGGCCCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTGGTT 663
Db 865 AAAGATCCCAAGCGGAGATGACCATTTGCCAGAGCCTGGAACATTCCTGGAT 917

RESULT 11

US-09-186-277-3

; Sequence 3, Application US/09186277

; Patent No. 6171841

; GENERAL INFORMATION:

; APPLICANT: AKIRA, SHIZUO

; APPLICANT: KAWAI, TARO

; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE

; FILE REFERENCE: 081356/0128

; CURRENT APPLICATION NUMBER: US/09/186,277

; CURRENT FILING DATE: 1998-11-05

; EARLIER APPLICATION NUMBER: JP97/261589

; EARLIER FILING DATE: 1997-09-26

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 2132

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (94)..(1455)

; US-09-186-277-3

Query Match 5.6%; Score 168.6; DB 3; Length 2132;
Best Local Similarity 54.8%; Pred. No. 3e-26;
Matches 358; Conservative 0; Mismatches 289; Indels 6; Gaps 1;

Qy 17 CGGCCCGCAGGATACAGGAGCGAGAGATCTGTCGCGCGGTGAGCCACCCGCTGGTCAAG 76
Db 265 CGGAGGAGATCGAGCGGAGGTGAACATCTGCGGGAGATCCGGCACCCCAACATCATC 324
Qy 77 GGGCTGTGGACCAAGTTTGAGACCCGCAAGACCTCATCTCTCGGAGCTGTGCTCA 136
Db 325 ACCGTGACGACATCTTCGAGAAACAAGCAGACGTGGTCTCTATCTCTGGAGCTGTCT 384
Qy 137 TCCGAGGAGTGTCTGGACCGCTGTACAGGAAGGCGTGTGTGACGAGGCGGAGGTCAAG 196
Db 385 GCGGGGAGCTCTTTGACTTCTTGGCGAGAAAGAGTCGCTGACGAGGAGCGAGGCCACC 444
Qy 197 GTCTACATCCAGCAGCTGTGGGGGCTGCACTACCTGCACAGCCATGGCGTTCTCCAC 256

Db 445 CAGTTCCTCAAGCAGATCCTGGACGGGTTCACTACCTGCACTCTAAGCCGATCGCAC 504
Qy 257 CTGACATAAAGCCCTCTAAACATCCTGATGGTGATCCTGCGCGGAAGAC-----ATT 310
Db 505 TTTGACCTGAAGCCGGAACATCATCTGCTGACAAAGACGTGCCAACCCACGAATC 564
Qy 311 AAAATCTGGCAGCTTTGGCTTTGCCAGAAATCATCACCCAGCAGAGCTGCAAGTTCAGCCAG 370
Db 565 AAGCTCATCGACTTCGGCATCGCGACAAGATCGAGGCGGGAACAGATTCAAGAAACATC 624
Qy 371 TAGGGCTCCCTGAGTTGCTCTCCCGAGATCATCCAGAGACCTGTGAGGGAAGCC 430
Db 625 TTCGGCACCCCGGAGTTTGGCCCCAGAGATTGTGAATATGAGCCGCTGGGCGCTGGAG 684
Qy 431 TCCGACATTTGGGCATAGGGTGTCTCTACCTCAGCCTGACCTGCTCATCCCCATT 490
Db 685 GCGGACATGTGGACATCGGTGTCTATACCTATATCTCTGAGCGGTGATCCCCGTTT 744
Qy 491 GCCGGCAGAGTGAACCTCTGAAACGTCTGAGGGGCGCGTGTATGAGGC 550
Db 745 CTGGCGAGACCAAGCAGGAGAGCTCACCAACATCTCAGCCGCTGAACTACGACTTCGAC 804
Qy 551 AGCCCCATGGTCCGACCTCAGCGAAGACGCCAAGACTTTCATCAAGGCTACGCTGCAG 610
Db 805 GAGGAGTACTTCAGCAACACCGAGCGAGCTGGCCAAAGGACTTCAATTCGCCGCGTCTCGTC 864
Qy 611 AGAGCCCTCAGGCCCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTGGTT 663
Db 865 AAAGATCCCAAGCGGAGATGACCATTTGCCAGAGCCTGGAACATTCCTGGAT 917

RESULT 12

US-09-949-016-5449

; Sequence 5449, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5449

; LENGTH: 1847

; TYPE: DNA

; ORGANISM: Human

; US-09-949-016-5449

Query Match 5.0%; Score 148.6; DB 4; Length 1847;
Best Local Similarity 51.5%; Pred. No. 5e-22;
Matches 368; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

Qy 34 GGAGCGAGACATCCTGGCGCGCTGAGCCACCCGCTGGTGCACGGGGCTGTGGACCAAGTT 93
Db 874 GGAGATTAGCATCATGAATGCTCTCCACCCCTCAAGCTGGTCCAGTGTGTGGATGCTTT 933
Qy 94 TGAGACCCCGCAAGACCTCATCTCTCTGAGCTGTGCTCATCCGAGGAGCTGTGGGA 153
Db 934 TGAAGAAAAGGCCAAACATGTCATGTCTCTGGAGATCGTGTGAGGAGGGAGCTGTTGA 993
Qy 154 CCG-...CCTGTACAGGAAGGCGTGTGACGAGGCGGAGGTCAAGTCTTACATCCAGCA 210
Db 994 GCGCATATTGACGAGGACTTTTGAGCTGACGGAGCGGTGATGATCAAGTACATGCGCA 1053

Qy	211	GCTGGTGGAGGGCTGCACCTACCTGCAACAGCCATGGCGTTCTCCACTGGACATATAAGCC	270
Db	1054	GATCTCGGAGGGAGTGGAGTACATCCACAAGACAGGCATCGTGACCTGACCTCAAGCC	1113
Qy	271	CTCTAACATCCTGATGCTGCATCCTGCGCCGGGAAGACATTAAATCTGCGACATTTGGCTT	330
Db	1114	GGAGAACATCATGTGTCTCAACAAAGACGGGCAACAGGATCAAGCTCATCGACTTTGGTCT	1173
Qy	331	TGCCCAAGAACATCACCCAGCAGAGCTGCAGTTTCAGGCCTGAGTTTCAGTTCGT	390
Db	1174	GGCCAGAGGCTGGAGAACGGGGGTCTCTGAAGTCTCTTTTGGACACCCAGATTTGT	1233
Qy	391	CTCCCCGAGATCATCCAGCAGAACCTGTGAGCGAAGCCTCCGACATTTGGGCCATGGG	450
Db	1234	GGCTCTGGAAGTGATCAACTATGAGCCCATCGGCTACGCCACACACATGTGGAGCATCG	1293
Qy	451	TGTCACTCTCACTGAGCTGACCTGCTCATATCCCAATTTGCGCGGAGAGTAGACCGTGC	510
Db	1294	GGTCATCTGCTATCATCTAGTCACTGAGTGCCTTTCCCTTTCATGGGACAAACGATACGA	1353
Qy	511	CACCTCTCAACGTCTCTGGAGGGGCGGTGTCATGGAGCAGCCCCATGCTGCCCACT	570
Db	1354	AACCTGGCCAAAGCTTACCTCAGCCACTGGGACTTCGACGAGAGCATTCGATGAGAT	1413
Qy	571	CAGCGAAGACGCCAAAGACTTCATCAAGGCTACGCTGCAGAGAGCCCTCAGGCCCGGCC	630
Db	1414	CTCGGACGATGCCAAGATTTCATCAGCAATCTGCTGAAGAAGAGATATGAAAACCGCCT	1473
Qy	631	TAGTGGCGGCCAGTGCCTCTCCACCCCTCGTTCCTGAAATCCATGCTCGGAGGAGGC	690
Db	1474	GGACTGCGCAGTGCCTTCAGCATCATGGCTAAATGAAGATACCAAGAACATGGAGGC	1533
Qy	691	CCACTTCATCAACACCAAGCAGCTCAAGTTCTCTTGGCCCGAAGTCGCTGGCAG	745
Db	1534	CAAGAAACTCTCAAGACCGGATGAAGAAGTACATGCGAAGAAAGGAAATGGCAG	1588

RESULT 13

```

US-09-917-254-41
; Sequence 41, Application US/09917254
; Patent No. 6703204
; GENERAL INFORMATION:
; APPLICANT: Mutter, George
; APPLICANT: Baak, Jan
; TITLE OF INVENTION: Prognostic Classification of Breast Cancer
; FILE REFERENCE: B0801/7324(JRV)
; CURRENT APPLICATION NUMBER: US/09/917,254
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/222,093
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
; LENGTH: 5926
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-917-254-41

```

Qy	211	GCTGTGTGGAGGCGCTGCACTACCTGACAGCCCATGGCTTCTCCACTGGACATAAAGCC	270
Db	4823	GATCTCGAGGGAGTGGAGTACATCCACAAGCAGGCGCATCGTGCACCTGGACCTCAAGCC	4882
Qy	271	CTCTAAACATCCTGATGGTGTGATCCTCTCGCGGGAAGACATTAAAAATCTGCGACATTTGGCTT	330
Db	4883	GGAGAACATCATGTGTCTCAACAAGACGGGCACCAGGATCAAGCTCATCGACTTTGGTCT	4942
Qy	331	TGCCCAAGAACATCACCCGACGAGCTGCAAGTTTCAGCCAGTACGGCTCCCTCGAGTTCGT	390
Db	4943	GGCCAGGAGGCTGGAGAAATGCGGGGTCTCTGAAGTTCCTCTTTGGCACCCCAAAATTTGT	5002
Qy	391	CTCCCCGAGATCATCCAGCAGAAACCTGTGACGGAAGCCTCCGACATTTGGGCCCATGGG	450
Db	5003	GGCTCTGAAGTGATCAACTATGAGCCCATCGGCTACGCCACAGACATGTGGAGCATCGG	5062
Qy	451	TGTCACTCTCTACTCAGCGCTGACCTGCTCATCCCCATTTGCGCGGAGAGTGACCGTGTC	510
Db	5063	GGTCATCTGTACATCTCTAGTCAGTGGCCCTTTCCGCCCTTCATGGGAGACAACGATAAACGA	5122
Qy	511	CACCTCTCTGAACGTCTCTGAGGGGGCGGTGTCATGGAGCAGGCCCATGGCTGCCACACT	570
Db	5123	AACTTTGGCCAAAGTTTACCTCAGCCACTTGGGACTTCGAGGACAGGCATTCGATGAGAT	5182
Qy	571	CAGCGAAGACGCCAAAGACTTCATCAAGGCTACGCTGCAGAGAGCCCTTCAGGCCCGGCC	630
Db	5183	CTCCGACGATGCCAAGGATTTTCATCAGCAATCTGCTGAAGAAAGATATGAAAAACCGCCT	5242
Qy	631	TAGTGGGGCCAGTGCTCTCCACCCCTGGTTCTCTGAAATCCATGCCTCGGAGGAGGCC	690
Db	5243	GGACTGCAGCGATGCTTCCTCAGCATCCATGGCTTAATGAAAGATACCAAGAACATGGAGGC	5302
Qy	691	CCACTTTCATCAACACCAAGCAGCTCAAGTTCTCTCTGGCCCGAAGTCGCTGGCGAG	745
Db	5303	CAGAAGAACTCTCCAGGACCGGATGAGAAGTACATGSCAAGAAAGAAATATGGCAG	5357

RESULT 14

```

US-09-797-039--9
; Sequence 9, Application US/09797039
; Patent No. 6730491
; GENERAL INFORMATION:
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 2504, 15977, AND 14760, NOVEL PROTEIN
; TITLE OF INVENTION: KINASE FAMILY MEMBERS AND USES THEREFOR
; FILE REFERENCE: 10448-017001
; CURRENT APPLICATION NUMBER: US/09/797,039
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,061
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-797-039--9

```

```
Db 1107 GAGGATTGTGATGAGGACTACCATCTGACCGAGGTGGACACCATGTGTGTTGTCAGGCA 1166
Qy 211 GCTGGTGGAGGGCTGACCTACCTGCGACAGCCATGGGGTCTTCCACCTGGACATAAAGCC 270
Db 1167 GATCTGTGACGGGATCCTCTTTCATGSCAAAGATGAGGGTTTTGACCTGGACCTCAAGCC 1226
Qy 271 CTCTAAACATCTGATGTGTCATCTGCGCGGAAGACATTAAATCTGCGACTTTGGCTT 330
Db 1227 AGAGAACATCTGTGTGTCAACACACCGGGCATTTGTGAGATCAATTGACTTTGGCCT 1286
Qy 331 TGCCGAGAACATACCCAGCAGAGCTGCAAGTTGAGCCAGTACGGCTCCCTCGAGTTTCGT 390
Db 1287 GGCACGGAGGTATTAACCCCAACGAGAGAGCTGAAGGTGAACCTTTGGGACCCCGAGAGTTCT 1346
Qy 391 CTCCTCCGAGATCATCAGCAGAACCTCTGTGAGGAAGCCCTCCACATTTGGGCCATGGG 450
Db 1347 GTCACTCTGAGGTGTGAATTATGACCAATCTCCGATAAGACAGACATGTGGAGTATGGG 1406
Qy 451 TGTCACTCTCTACCTGAGCCTGACCTGCTCATCTCCCAATTTGCGCGGAGAGTGACCGTGC 510
Db 1407 GGTGATACCTTACATGTCTGAGGGGCTCTCCCTTCTCTGGAGATGATGACAGAGA 1466
Qy 511 CACCTCTCTGAACGTCTGGAGGGCGGCTGTATGAGCAGAGCCCTATGGCTGCCCACT 570
Db 1467 GACCTTAAACAACTTCTATCTGGCAACTGTGTACTTTGATGAAGAGACCTTTTGAGGGCGT 1526
Qy 571 CAGCGAAGACCCAAAGACTTCATCAAGGCTACGCTGACAGAGCCCTCAGGCGCCGCC 630
Db 1527 ATCAGACGAGGCCAAAGACTTTGTCTCAACCTCATCGTCAAGGACGAGGGGCCCGGAT 1586
Qy 631 TAGTGGCGCCAGTGCCTCTCCCACTCTGCTTCTCTGAA 669
Db 1587 GAACGCTGCCAGTGTCTGCGCCATCTCTGCTCAACAA 1625
```

RESULT 15

```
US-09-797-039-7
; Sequence 7, Application US/09797039
; Patent No. 6730491
; GENERAL INFORMATION:
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 2504, 15977, AND 14760, NOVEL PROTEIN
; TITLE OF INVENTION: KINASE FAMILY MEMBERS AND USES THEREFOR
; FILE REFERENCE: 10448-017001
; CURRENT APPLICATION NUMBER: US/09/797,039
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,061
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2046
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (119)...(1906)
; NAME/KEY: misc_feature
; LOCATION: (1)...(2046)
; OTHER INFORMATION: n = A,T,C or G
US-09-797-039-7
```

```
Query Match 4.8%; Score 143; DB 4; Length 2046;
Best Local Similarity 52.6%; Pred. No. 7.9e-21;
Matches 336; Conservative 0; Mismatches 300; Indels 3; Gaps 1;

Qy 34 GGACGGAGACATCTCTGGCGGGCTGAGCCACCGCTGGTCACTGGGGCTGTGGACCAAGTT 93
Db 1105 GGAGATTGAGGTGATGAACCAAGCTGACCAACCGCAATCTGATCCAGCTGTATGAGCCAT 1164
Qy 94 TGAGACCGGCAAGACCCCTCATCTCTCTGGAGCTGTGCTCATCCGAGGAGCTGTGCTGGA 153
```

```
Db 1165 CGAGACTCCGCATGAGATCGTCTCTGTTTCATGTGAGTACATCGAGGGCGGAGAGCTCTTCGA 1224
Qy 154 CCG---CCTGTACAGGAAGGGCGTGTGACGGAGGGCGAGGTCAAGGTCTTACATCCAGCA 210
Db 1225 GAGGATTTGTGATGAGGACTTACCATCTGACGGAGGTGGACACCATGGTGTGTTGTCAAGCA 1284
Qy 211 GCTGGTGGAGGGGTGCACCTACCTGACACAGCCATGGCGTTCTCCACCTGGACATATAAGCC 270
Db 1285 GATCTGTGAGGGATCCTCTTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCC 1344
Qy 271 CTCTAACATCTGATGTGTCATCTGCTCCCGGGAAGACATTAAATCTGCGACTTTGGCTT 330
Db 1345 AGAGAACATCTGTGTGTCAACACCCACCGGGCATTTGGTGAAGATCAATTGACTTTGGCCT 1404
Qy 331 TGCCGAGAACATCACCCAGCAGAGCTGCAAGTTTTCAGCCAGTACGGCTCCCTCGAGTTTCGT 390
Db 1405 GGCACGGAGGTATAACCCCAACGAGAGCTGAAGGTGAACCTTTGGGACCCCGAGAGTTCTCT 1464
Qy 391 CTCCTCCGAGATCATCCAGCAGAACCTCTGTGAGCGAAGCCTCCGACATTTGGGCGCATGGG 450
Db 1465 GTCACTGAGGTGGTGAATTTATGACCAATCTCCGATAAGACAGACATGTGGAGTATGGG 1524
Qy 451 TGTCACTCTCTTACCTCAGCCTGACCTGCTCATCTCCCAATTTGCGGCGGAGAGTGACCGTGC 510
Db 1525 GGTGATCACCTACATGCTGTGAGCGGCTCTCCCTTCTCTGGGAGATGATGACACAGA 1584
Qy 511 CACCTCTCTGAACGTCTCTGGAGGGCGGCTGTCTATGGAGCAGCCCATGGCTGCCACCT 570
Db 1585 GACCTTAAACAAAGCTTCTATCTGGCACTGGTACTTTGTATGAAGAGACCTTTGAGGGCGT 1644
Qy 571 CAGCGAAGACGCCAAAGACTTCATCAAGGCTAGGCTGACAGAGAGCCCTCAGGCGCCGCC 630
Db 1645 ATCAGACGAGGCCAAAGACTTTGTCTCAACCTCATCGTCAAGGACGAGGGGCCCGGAT 1704
Qy 631 TAGTGGCGCCAGTGCCTCTCCCACTCTGCTTCTCTGAA 669
Db 1705 GAACGCTGCCAGTGTCTGCGCCATCTCTGCTCAACAA 1743
```

Search completed: March 21, 2005, 14:39:39
Job time : 533.683 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:02:07 ; Search time 193.228 Seconds
(without alignments)
9323.380 Million cell updates/sec

Title: US-10-077-130-6_COPY_13600_14700
Perfect score: 1101
Sequence: 1 cacagcagcacactgtgac.....agcaggaggcccatgttc 1101

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45.8	4.2	1350	4	US-09-902-540-5939
C 2	45.8	4.2	1351	4	US-09-902-540-192
C 3	45	4.1	2112	4	US-09-902-540-3277
4	45	4.1	17480	4	US-09-902-540-1151
5	45	4.1	33529	3	US-09-144-085-3
C 6	44.8	4.1	765	4	US-09-252-991A-15050
C 7	44.8	4.1	1329	4	US-09-252-991A-14932
C 8	44.8	4.1	5121	4	US-09-252-991A-15189
9	44	4.0	735	4	US-09-252-991A-14239
C 10	44	4.0	1410	4	US-09-252-991A-14025
C 11	43.8	4.0	1811	4	US-09-902-540-375
C 12	42.6	3.9	1017	4	US-09-252-991A-14630
C 13	42.6	3.9	2502	4	US-09-252-991A-14994
C 14	42.6	3.9	2505	4	US-09-252-991A-14750
C 15	42	3.8	3397	4	US-09-949-016-2188
C 16	42	3.8	119801	4	US-09-949-016-13453
C 17	41.8	3.8	4403765	3	US-09-103-840A-2
C 18	41.8	3.8	4411529	3	US-09-103-840A-1
C 19	41.4	3.8	575	4	US-09-270-767-3998
C 20	41.4	3.8	575	4	US-09-270-767-19280
C 21	41.4	3.8	2010	4	US-09-902-540-9547
C 22	41.4	3.8	14807	4	US-09-902-540-1085
C 23	41.2	3.7	1926	3	US-09-249-585A-2
24	41.2	3.7	1926	3	US-09-410-399-3
25	41.2	3.7	2580	3	US-09-050-863-2
26	41.2	3.7	2580	3	US-09-359-081-2
C 27	41.2	3.7	5452	2	US-09-130-114-1

C 28	41.2	3.7	8705	4	US-09-647-344A-14	Sequence 14, Appl
29	41.2	3.7	9600	3	US-08-910-647-1	Sequence 1, Appl
30	41.2	3.7	9600	3	US-09-620-925-1	Sequence 1, Appl
31	41.2	3.7	10596	1	US-07-884-811-15	Sequence 15, Appl
32	41.2	3.7	10596	1	US-07-885-971-15	Sequence 15, Appl
33	41.2	3.7	10596	1	US-08-087-783A-15	Sequence 15, Appl
34	41.2	3.7	10596	1	US-08-194-088B-15	Sequence 15, Appl
35	41.2	3.7	10596	2	US-08-194-087-15	Sequence 15, Appl
36	41.2	3.7	10596	5	PCT-US93-04648-15	Sequence 15, Appl
C 37	41.2	3.7	16080	4	US-09-724-566A-48	Sequence 48, Appl
C 38	41.2	3.7	16080	4	US-09-471-669A-48	Sequence 48, Appl
C 39	41	3.7	1065	4	US-09-976-594-833	Sequence 833, App
40	41	3.7	4078	4	US-09-016-434-1120	Sequence 1120, Ap
41	41	3.7	5926	4	US-09-917-254-41	Sequence 41, Appl
C 42	40.8	3.7	843	4	US-09-724-797-33	Sequence 33, Appl
43	40.8	3.7	1023	4	US-09-902-540-3025	Sequence 3025, Ap
44	40.8	3.7	3073	4	US-09-902-540-643	Sequence 643, App
45	40.4	3.7	486	4	US-09-252-991A-4065	Sequence 4065, Ap

ALIGNMENTS

RESULT 1

US-09-902-540-5939
; Sequence 5939, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5939
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-5939

Query Match	4.2%	Score 45.8;	DB 4;	Length 1350;
Best Local Similarity	45.6%	Pred. No. 0.28;		
Matches 161;	Conservative 0;	Mismatches 192;	Indels 0;	Gaps 0;
Qy	382	GAGCGCATCCAGCCGCGGTTCGAGGTGTTCCAGGGTTCGCCAAGGATGCTG	441	
Db	361	GAAGCGGTGCGCGCGCGGACAGGCGCGCGGTTCGCCAAGGATGCTG	420	
Qy	442	GTGATCAAGGCGTTTCAGCGGAGAGACCCAGGCGGAGTACCACTGTGGCTTGCTCAGGCG	501	
Db	421	TCCACCATCTCAAGGGGTGAGACCTGGGAGCAACGTGCGGAGGTGTCGCGCGCC	480	
Qy	502	TCCATCTGCCCTGGCGGTGCCACCTTCCAGTGGCACTAGAGCCAGCCTCTGTGGATGAG	561	
Db	481	TCCAGCGAGCAGGTGCGAGGCCACGCGCGCTGGTCCAGGGCACCGCGAAGGTGAGCGAA	540	
Qy	562	CCCCCTAGCCAGCTTCCCCCGAGGACCCAGGAGGTGACCTGCACCTACTGTGG	621	
Db	541	CAGGAGCGGCGCATCGCCACATCCGCGCGCGAGGACAGGCGCGCGCTGGCG	600	
Qy	622	GAGCGCTTGGCTCGGAAACGTTCGATGAGCGGTGAGCCACGCTGGACTTCCATTAGCGAG	681	
Db	601	CAGGCGCGCGGAGATCGCGCGGATGTCGCGGAGACGATGAGGCCACCTCCGAGCAG	660	
Qy	682	CTGCGAGGAGGACGCGCGCTCGCAGCGCTGCCACAGGAGGACAGAGGT	734	
Db	661	CGCGCGCGCTGCGGACGCTGGTGGTCTCCAAACGCGGACGCTCGCGGAGGCGGT	713	

RESULT 2
US-09-902-540-192/c
; Sequence 192, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 192
; LENGTH: 1351
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-192

Query Match 4.2%; Score 45.8; DB 4; Length 1351;
Best Local Similarity 45.6%; Pred. No. 0.28;
Matches 161; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
QY 382 GAGCGCATCCAGCCCGGTGGCGGTTCCGAGGTGTTCCAGGTCGCGCAACAGATGCTG 441
DB 991 GAAAGCGTGGCGGCGCGAGAGAGGCGCGCTGGCAGGGATGCGGAGCGCGCTG 932
QY 442 GTGATCAAGGGTTCACGCGCAAGAACAGAGGCGAGTACCACTGTGGCTGCTCAGGGC 501
DB 931 TCCACCATCTCAAGGGGTGGAGACTGTGGGACACAGTGGCGAGGTGTGGCGGCC 872
QY 502 TCCATCTGCTCGGCTGCCACCTTCAGGTGGCACTGAGCCAGCCCTCTGTGGATGAG 561
DB 871 TCCAGCGAGCAGTGCAGGCCACGAGCGCTGTCCAGGGCACCAGGAGGTGAGCGAA 812
QY 562 GCCCTCAGCCAGCTTCCCGCCGAGGAGCGCCAGGAGGAGTGCCTCAGCTACTGTGG 621
DB 811 CAGGAGCGGGCATCCGACATCCCGCGCCGAGCAGGAGGACAGCGCGCGCTGGCG 752
QY 622 GAGGCGCTGGCTCGAAACGTGCGATGAGCGGTGAGCCACGCTGGACTCCATTAGCGAG 681
DB 751 CAGGCGCGCGGAGATGCGGCGCATGTCGGCGAGAGATGCAGGCCACCTCCGAGCAG 692
QY 682 CTGCAGAGGAGCGCGCTCGCAGCGCTGCCACAGGAGGAGGAGGT 734
DB 691 GCGCGCGCGCTCGCGACGTGGTGGCTCCAAACGGGCGAGCTCGCGGAGGCGGT 639

RESULT 3
US-09-902-540-3277/c
; Sequence 3277, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3277
; LENGTH: 2112
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3277

Query Match 4.1%; Score 45; DB 4; Length 2112;
Best Local Similarity 46.7%; Pred. No. 0.48;
Matches 178; Conservative 0; Mismatches 200; Indels 3; Gaps 1;
QY 217 GAACCGGTTACCTGCCCCCAGACAGTGGCGGTTTCAGAGCCACCGAGCCTGTGCTCCC 276
DB 1494 GAAGGGTAAACCCCGCCCCCAGCACCAGCGGATGATGGCGATGACCGCGCGCGAGGGC 1435
QY 277 CAGCCTCAGCCCTGAGAGCCGCGAGGTGGCAGCTGGTGAAGATGTCTTCTGGAGCTT 336
DB 1434 GAACGGGTCCCGCCCGCCGTTAGCGGCTTCCCGATGCGGAAGATGCCGACAGCGAGTC 1375
QY 337 GAGGTGGTGGCTGAGGCTGGTGGTCACTTGGGCACAAGGAATGGAGCGCATCCAGCCC 396
DB 1374 GAGCCTGTCCGCGATACCGATGAGCGCGCGCGGTCTTGGCTGGGCGACGCGCTCCTCGGC 1315
QY 397 GGTGGCGGTTCCGAGGTGGTCTCCAGGGTCCGCAACAGATGCTGGTGATCAAGGGCTTC 456
DB 1314 GCCAGCGGCGAGGTAGTCTCCGCGATGGCCAGAGCCACGGCGTCCGGCTCAGCGCGC 1255
QY 457 ACGGCAGAAGACACAGGGCGAGTACCACTGTGGCTGGCTCAGGGCTCAATCTGCCCTGGC 516
DB 1254 GCGGGCGTACTCCCGCGCCCATGATGCCCTGGAGCTCGGGGAACCTCGGCCA---CCATGCC 1198
QY 517 GCTGCCACCTTCCAGGTGGCACTGAGCCCGACCTCTGTGGATGAGGCCCTCAGCCCGAGC 576
DB 1197 GGTGACGAGGTGGCGCTTGGCGAGGTGGCGCGCGCTCGATGTTAGCCGCTCCTCCCTGC 1138
QY 577 TTGCCCCCGAGGCGAGCCAG 597
DB 1137 TCGCCCCGCTCTCTGGGCCAG 1117

RESULT 4
US-09-902-540-1151
; Sequence 1151, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1151
; LENGTH: 17480
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1151

Query Match 4.1%; Score 45; DB 4; Length 17480;
Best Local Similarity 46.7%; Pred. No. 0.79;
Matches 178; Conservative 0; Mismatches 200; Indels 3; Gaps 1;
QY 217 GAACCGGTTACCTGCCCCCAGACAGTGGCGGTTTCAGAGCCACCGAGCCTGTGCTCCC 276
DB 7398 GAAGGGTAAACCCCGCCCCCAGCACCAGCGGATGATGGCGATGACCGCGCGCGAGGGC 7457
QY 277 CAGCCTCAGCCCTGAGAGCCGCGAGGTGGCAGCTGGTGAAGATGTCTTCTGGAGCTT 336
DB 7458 GAACGGGTCCGCGCGCCCGCTTGGCGATGGCGGAAGATGCCGACAGCGAGTC 7517
QY 337 GAGGTGGTGGCTGAGGCTGGTGGTCACTTGGCACAAGGAATGGAGCGCATCCAGCCC 396
DB 7518 GAGCCTGTCCGCGATACCGATGAGCGCGCGCGGTCTTGGCTGGGAGCGCGCTCCTCGGC 7577
QY 397 GGTGGCGGTTTCGAGGTGGTCTCCAGGGTTCGCAACAGATGCTGGTGATCAAGGGCTTC 456

Db 7578 GCCACGGCGCAGGTAGTCTCCGCGATGCGCAGAGCCAGCGCGTTCGGCTTACCCGCCGC 7637
Qy 457 ACGGAGAGACACAGGCGAGTACCACTGTGCGCTTGCTTCAAGGCTTCCATCTGCGCTCGG 516
Db 7638 GCGGCGGTACTCCGCGCCATGATGCCCTGGAGCTCGGGAACCTCGCCCA---CCATGCC 7694
Qy 517 GCTGCCACTTCCAGGTGCACTGAGCCAGCCTCTGTGGATGAGGCGCCCTCAGGCCAGC 576
Db 7695 GGTGACGAGGTGCGCCTTGGCCAGAGGTGCGCGCGCTCGATGTTAGCGCGCTCCCGCTGC 7754
Qy 577 TTGCCCCCGAGGACGCCAG 597
Db 7755 TCGCCCCCGCTCTCGGCCAG 7775

RESULT 5

US-09-144-085-3

; Sequence 3, Application US/09144085

; Patent No. 628099

; GENERAL INFORMATION:

; APPLICANT: Gustafsson, Claes

; APPLICANT: Betlach, Mary C.

; APPLICANT: Ashley, Gary

; APPLICANT: Julien, Bryan

; APPLICANT: Ziermann, Rainer

; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA

; FILE OF INVENTION: THEREFOR

; FILE REFERENCE: 30062-20020.20

; CURRENT APPLICATION NUMBER: US/09/144,085

; EARLIER FILING DATE: 1998-08-31

; EARLIER APPLICATION NUMBER: 09/010,809

; EARLIER FILING DATE: 1998-01-22

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 33529

; TYPE: DNA

; ORGANISM: Sorangium cellulosum

US-09-144-085-3

Query Match 4.1%; Score 45; DB 3; Length 33529;

Best Local Similarity 44.8%; Pred. No. 0.92;

Matches 298; Conservative 0; Mismatches 360; Indels 7; Gaps 3;

Qy 74 CGCTGGAGGTGAAGAGGGGGCCACAGCCAGTGGCGGCTGTGCCAGAGCTGTGTGCGCTG 133
Db 13643 GCGCGTGGCGCGCGCCAGAGACGCGCTCGAGGACCTGGCGCGCGCGCTGTGG 13702
Qy 134 GACCGAGTGTGTGTGATGGCTTGGCCCGCGGGGAGACCTACCGCTTCCGTGTGGCAG 193
Db 13703 GCGTGTCCGCGCGCGCGAAGACGAGACACCCCGAAGCGGGGCTGCGCTTGATGATGTGG 13762
Qy 194 CTGTGGCGCTGTGGGTGTGGGAACCGGTTTCACTGCCCGCAGACAGTGGCGCTTGAG 253
Db 13763 GGACGACCGCTGAGACGTGGGCTGTGGCGCGCGCGCTGTGGCGCAGAGCGCGG 13822
Qy 254 AGCCACCGAAGCTGTGTGCTTCCCCAGCCCTCAGCCCTCAGAGCGCGGAGGTGGCAGCTG 313
Db 13823 AGCTTGCCTGCGCGGGCGCGCTGCGCTGGCGCGCGCGCTGTGGCGCAGAGCGGTAG 13882
Qy 314 GTGAAGATGTCTCTCTGAGCTTGAAGTGTGGCTGTGAGGCTGTGAGTCACTTGGCAAC 373
Db 13883 CGGAAGAGCTCACCGGACCGCGAGCTGGACCTTCGCGGACGCTGTGTGTGACCGCG 13942
Qy 374 AGGGAATGAGCGCATCCAGCCGCTGGCGGCTTCACTGAGTGTCTCCAGGCTCGGCAAC 433
Db 13943 GGACAGGGAGCTGGGTCAGGCGGTTCGCGCGCACTGTGTGGCGCGCAGCGGGTGC 14002
Qy 434 AGATGCTGGTGA--TCAAGGGCTTCAAGGAGAGACAGGCGCAGTACCACTGTGGCT 491
Db 14003 ACCTGTGCTGACGTGCGCGCGCGGCTGGAGGCGCGCGCGCGCGAGCTTGTGGAT 14062

Qy 492 GGCTCAGGGTCCATCTGCGCTGCCACTTCCAGGTGGCACTGAGCCCGCCTC 551
Db 14063 CCGTCGCGAGCTCGCGCCGAGACGGTGAACGGTGGCGCGCTCGACGTTCGAGC--G 14120
Qy 552 TGTGATCAGGCGCCCTCAGCCAGCTTGCCTCCCGCAGGCGAGCCAGAGGGTGAACCTGCA 611
Db 14121 GGAGAGGTTCGCGCTGTGCTGGCGGATCGACGCGCGCGCGCTGAGCGCGGTGT 14180
Qy 612 CTTACTTGGGAGCGCTGGCTCGGAAACGTGCGATGAGCGGTGAGCCACCTGGAATC 671
Db 14181 GCACCTGCGCGCGCTCGACGCGGCTGCTCGCGCGCAGAGCGCGAGCGCTCTC 14240
Qy 672 CATTAGCAGCTGCAGAGGAGCGCGCTCGCAGCGCTGCGCAGGAGCGAGAGA 731
Db 14241 --GCGGTGCTGGCGCGAAGGTGACGCGGCGCTGCACCTGCACGAGCTGACCGGGA 14297
Qy 732 GGTGG 736
Db 14298 GGTGG 14302

RESULT 6

US-09-252-991A-15050/c

; Sequence 15050, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; EARLIER FILING DATE: 1999-02-18

; EARLIER APPLICATION NUMBER: US 60/074,788

; EARLIER FILING DATE: 1998-02-18

; EARLIER APPLICATION NUMBER: US 60/094,190

; EARLIER FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 15050

; LENGTH: 765

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-15050

Query Match 4.1%; Score 44.8; DB 4; Length 765;

Best Local Similarity 44.0%; Pred. No. 0.42;

Matches 190; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

Qy 46 GATGAGCGGTGTCTCTGTGGCTACCGCTGAGGTGAAGAGGGGGCCACAGCCAG 105
Db 641 GCTGTGCGGTGTGAGGGCGACGCGCGCAGTTTCAGGCGCGCTTGTGCGGCTT 582
Qy 106 TGGCGGCTGTGCCACGAGCTGTGCTGGACCCGAGTGTGTGGATGGCTTGGCCCCC 165
Db 591 GCCCGCAGTTCAGGCGGTAGCCGGGGTGGCGGTTTGGCAGAGACGACCTCGCGT 522
Qy 166 GGGAGAGCTTACCGCTTCCGTGTGGCAGCTGTGGGCCCTGTGGGTGTGAGGAAACGGTT 225
Db 521 CTGCGCCAGCGGTGGCCATCGCTTCGCGACTTCGTAGCGAGGCTGGCGAGGCGCTG 462
Qy 226 CACCTGCCCCAGACAGTGGCGCTTCGAGAGCCACCGAAGCTGTGCTCCCGAGCCCTCA 285
Db 461 GAGCAGCGCTTGGCGGCGGCGAGGCGCGCTGTTGCTGGAGGTGGCGGTGAGGATGGC 402
Qy 286 GCCCTGAGAGCGCGCAGGTGGCAGTGTGTAAGATGTCTCTCTGAGCTTGAAGTGGTG 345
Db 401 GTTGATCGCTCGGTGAGTTGCGGAGTTGCTGGGATCTCTGTGCGGCTTGCAGCGGCG 342
Qy 346 GCTGAGGCTGTGAGGTCACTGCGCAAGGGAATGGAGCGCATCCAGCCCGGTGGCGG 405
Db 341 GCGCGCTGGAGAGTTCGCGGTTCGCGCGCGCGAGGGCGGCGACTTCGCTGCGGAG 282
Qy 406 TTCAGGTGTCTCCAGGTCGCGCAACAGATGCTGGTGTATCAAGGCTTTCACGCGAGNA 465

Db 281 GTCTTCAGGTGTTCCAGCGCTGGCGTTGCTGCTGGTGTTCGCGGGCGGCGCTGGGCGAG 222
QY 466 GACCAGGGCGAG 477
Db 221 GTCGAGGACCAG 210

RESULT 7

US-09-252-991A-14932/c
; Sequence 14932, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14932
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14932

Query Match 4.1%; Score 44.8; DB 4; Length 1329;
Best Local Similarity 44.0%; Pred. No. 0.47;
Matches 190; Conservative 0; Mismatches 242; Indels 0; Gaps 0;
QY 46 GATGAGGCGGTGCTCTGTGGCTACCGCTGAGGCTGAAGGAGGGGGCCACAGGCCAG 105
Db 1146 GCTGTGCGGTTGGCTTGTAGGGCCAGCGCGCGAGTTGCGAGCGCGCTTGTGCGCCTT 1087
QY 106 TGGCGGCTGTGCCACAGCTGTGTGCTGACCCGAGTGTGTGTGATGTGCTGCCCCC 165
Db 1086 GCGCGCGACTTCCAGGCGGTAGCGGGGGTGGCGGTTTGGCAGGACGACCTGCCGGT 1027
QY 166 GGGGAGACTACCGTTCCTGTGTGACAGTGTGGCCCTGTGGTGTCTGGGGAACCGGTT 225
Db 1026 CTGCGCCAGCGGTGGCCATGCTTCGCGACTTCGTAGCCAGGCTGGCGAGCCCTG 967
QY 226 CACCTGCCCCAGAGTGGCTTGCAGAGCCAGGACCGAAGCCTGTGCTCCCGCCCTCA 285
Db 966 GAGCAGCGCTTGGCGGCGGCGAGGCGGCTGTGCTGGAGGTGGCGGTGAGATGGC 907
QY 286 GCGGCTGAGAGCCGCGAGTGGCAGCTGTGTGAAGATGTCTCTTGGAGCTTGGGTGTG 345
Db 906 GTTGATCGCTCGGTGAGTTGCGCGAGTTGCTGCGGATCTCTGTGCGGCTGGCAGGCGC 847
QY 346 GCTGAGGCTGTGAGTGTATCTGGCACAAGGATGAGCGCATCCAGCCGCTGGGGGG 405
Db 846 GCGCGCTGAGAGTGTTCGGGTGTTCGCGGCGCGCGAGCGCGACTTCGCTGGCGAG 787
QY 406 TTCGAGTGTGCTCCAGGGTGGCAACAGATGCTGTGATCAAGGGCTTCAACGGCAGAA 465
Db 786 GTCTTGACAGTGTTCAGCGCTGGCGTTGCTGTGTTGCGGGGCGGCTGGGCGAG 727
QY 466 GACCAGGGCGAG 477
Db 726 GTCGAGGACCAG 715

RESULT 8

US-09-252-991A-15189/c
; Sequence 15189, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15189
; LENGTH: 5121
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15189

Query Match 4.1%; Score 44.8; DB 4; Length 5121;
Best Local Similarity 44.0%; Pred. No. 0.66;
Matches 190; Conservative 0; Mismatches 242; Indels 0; Gaps 0;
QY 46 GATGAGGCGGTGCTCTGTGGCTACCGCTGAGGCTGAAGGAGGGGGCCACAGGCCAG 105
Db 4843 GCTGTGCGGTTGGCTTGTAGGGCGACGCGCGCAGTTGAGGCGCGCTTGTGCGCCTT 4784
QY 106 TGGCGGCTGTGCCACAGCTGTGTGCTGACCCGAGTGTGTGTGATGTGCTGCCCCC 165
Db 4783 GCGCGCGACTTCCAGGCGGTAGCGGGGGTGGCGGTTTGGCAGGAGCAGCCCTGCCGGT 4724
QY 166 GGGGAGACTACCGTTCCTGTGTGAGCAGTGTGGCCCTGTGGTGTCTGGGGAACCGGTT 225
Db 4723 CTGCGCCAGCGGTGGCCATGCTTCGCGGACTTCGTAGCCAGGCTGGCGAGCCCTG 4664
QY 226 CACCTGCCCCAGAGTGGCTTGCAGAGCCAGGACCGAAGCCTGTGCTCCCGCCCTCA 285
Db 4663 GAGCAGCGCTTGGCGGCGGCGAGGCGGCTGTGCTGGAGGTGGCGGTGAGATGGC 4604
QY 286 GCGGCTGAGAGCCGCGAGTGGCAGCTGTGTGAAGATGTCTCTTGGAGCTTGGAGTGTG 345
Db 4603 GTTGATCGCTCGGTGAGTTGCGCGAGTTGCGCGGCTGCTGTGCGGCTGGCAGGCGC 4544
QY 346 GCTGAGGCTGTGAGGCTATCTGGCACAAGGATGAGCGCATCCAGCCGCTGGGGGG 405
Db 4543 GCGCGCTGAGCAGTTCGCGGCTTTCGCGGCGCGCGAGGCGGCTTCGCTGGCGAG 4484
QY 406 TTCGAGTGTGCTCCAGGGTGGCAACAGATGCTGTGATCAAGGGCTTCAACGGCAGAA 465
Db 4483 GTCTTGACAGTGTTCAGGCGCTGGCGTTGCTGTGTTGCGGGGCGGCTGGGCGAG 4424
QY 466 GACCAGGGCGAG 477
Db 4423 GTCGAGGACCAG 4412

RESULT 9

US-09-252-991A-14239
; Sequence 14239, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14239
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14239

Query Match 4.0%; Score 44; DB 4; Length 735;
Best Local Similarity 53.5%; Pred. No. 0.63;
Matches 92; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 330 GGAGCTTCAGGTGGTGGCTGAGCTGGTGGATCATCTGGCAACAGGGAATGGAGCGCAT 389
DB 207 GGTGCCAGCGCTTCTGGACGATGCTGGTGGCTGACGCTGGCCACCTGCTGGATCAGT 266
QY 330 CCAGCCCGGTGGCGGTTCGAGGTGGTCTCCAGGGTGGCAACAGATGCTGGTGAATCAA 449
DB 267 TGAGCTCACAGAGGCGATCGACCGCTCTTCTCGGTGGCAACTGCTCCAGGTGCTCGC 326
QY 450 GGGCTTCACGCGCAGAGACCGAGCGAGTACCACTGTGGCCCTGGCTCAGGGC 501
DB 327 GGTATTCTAGCGAGGTGCGAATGCTGCGCAGCCGCTGATCAGGCC 378

RESULT 10
US-09-252-991A-14025/c
; Sequence 14025, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14025
; LENGTH: 1410
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14025

Query Match 4.0%; Score 44; DB 4; Length 1410;
Best Local Similarity 53.5%; Pred. No. 0.74;
Matches 92; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 330 GGAGCTTCAGGTGGTGGCTGAGCTGGTGGATCATCTGGCAACAGGGAATGGAGCGCAT 389
DB 1273 GGTGCCAGCGCTTCTGGACGATGCTGGTGGCTGACGCTGGCCACCTGCTGGATCAGT 1214
QY 330 CCAGCCCGGTGGCGGTTCGAGGTGGTCTCCAGGGTGGCAACAGATGCTGGTGAATCAA 449
DB 1213 TGAGCTCACAGAGGCGATCGACCGCTCTTCTCGGTGGCAACTGCTCCAGGTGCTCGC 1154
QY 450 GGGCTTCACGCGCAGAGACCGAGCGAGTACCACTGTGGCCCTGGCTCAGGGC 501
DB 1153 GGTATTCTAGCGAGGTGCGAATGCTGCGCAGCCGCTGATCAGGCC 1102

RESULT 11
US-09-252-991A-14025/c
; Sequence 375, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiesand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825

; SEQ ID NO 375
; LENGTH: 1811
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-375

Query Match 4.0%; Score 43.8; DB 4; Length 1811;
Best Local Similarity 44.8%; Pred. No. 0.87;
Matches 168; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
QY 148 GTGGATGGCCCTGGCCCGGGGAGACCTACCGCTTCCGCTGTGGCAGCTGTGGCCCTGTG 207
DB 678 GTGGAGGGAACGTTGTTGGGAGCGGCTTCCAGGCCCGCTATGGGCGGACGCGCAGATG 619
QY 208 GGTGCTGGGGAACCGGTTTACCTGCCCCAGACAGTGGCGCTTGCAGAGCCACCGAAGCCT 267
DB 618 GAGCGCTGGAGGTGGGCGACTCGCGCTTACCGTGGCCGCGCTTGGGAGCGGCTGGCG 559
QY 268 GTGCTCCCCAGCCCTCAGCCCTCAGAGCGGCGAGGTGGCAGCTGTGAGATCTCTCT 327
DB 558 GCGCCTCCCGAGCTTTTCGCGAGCGGACTGCGCCGTGGAGGGGCCCCGAGCGCGCTGTG 499
QY 328 CTGAGCTTGAGGTGGTGGCTGAGGCTGCTGAGGTGCTGTCACACAGGGAATGGAGCGC 387
DB 498 CTGAGCGCGCGCTCGAGCTTCCAGCGCGCGCGCGGCTGACCGCTGGATGCGGCG 439
QY 388 ATCCAGCCCGTGGCGGTTTCGAGGTGCTTCCAGGGTTCGCCAACAGATGCTGTGATC 447
DB 438 GCGCGCGCGGCTGGCTGCGGAGGTCCACGCGGCTTCCCGAGGAAGACGCGCGGTGT 379
QY 448 AAGGCTTTCAGCGCAGAGACCGGCGAGTACCACTGTGGCCCTGGCTCAGGCTCCATC 507
DB 378 GCGGACTGGAGCGGAGCGGCGAGGCGGCGGCTGCTGCGCACCGCTGGCG 319
QY 508 TCCCTCGCGCTGCC 522
DB 318 TTCGCGCGGCGGCC 304

RESULT 12
US-09-252-991A-14630/c
; Sequence 14630, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14630
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14630

Query Match 3.9%; Score 42.6; DB 4; Length 1017;
Best Local Similarity 51.3%; Pred. No. 1.4;
Matches 99; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 424 GGTGGCAACAGATGCTGATCAAGGGCTTTCAGGCGAGAACACGCGGAGTACCAC 483
DB 340 GCTCAGCTGGCGGCGCTGCTGCTCGGCTTTCGACGGAACACGCGGTGCTGAGCG 281
QY 484 TGTGGCTGGCTCAGGGCTCCATCTGCGCTGCGGCTGCCACCTTCCAGGTGGCTGAGC 543
DB 280 GGTGCCAGCTCCTCGGCGTCTTCCACCTTGGGTTGAAGCCAGCGAGTGGCGCACCC 221

Qy 544 CGAGCTCTGTGGATAGACGCCCTCTAGGCCACGTTGCCCCCGAGGAGCGACGCCAGGAGGT 603
Db 220 CATGGCCGCATCAGGCGGTAGGGCGGTCCATCTCGCCACGTCACGGCAGCGCGTT 161
Qy 604 GACCTGCACCTAC 616
Db 160 GGCGGCTTCCTTC 148

```

RESULT 13
US-09-252-991A-14994
; Sequence 14994, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.1136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14994
; LENGTH: 2502
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14994

```

Query Match	3.9%	Score 42.6;	DB 4;	Length 2502;
Best Local Similarity	51.3%;	Pred. No. 1.8;		
Matches 99;	Conservative 0;	Mismatches 94;	Indels 0;	Gaps 0;
QY	424	GGTCGGCAACAGATGCTGGTGATCAACGGGCTTCACGCGCAGAGACACGAGCGCGAGTACCAC	483	
Db	694	GCTCAGCTGCGCGCGCTCGTCTCGGGCTTTTCGACCGAACAGCGGGTCGGTCAGCGG	753	
QY	484	TGTGGCCTGGCTCAGGGCTCCATCTGCGCTCGGGCTGCCACCTTCAGGTGGCACTGAGC	543	
Db	754	CGTGCAGCCCTCTCGGGCTCTTCACTTTGCGGTTGAAGCCAGCCAGGTTGCGCACACC	813	
QY	544	CAAGCCTCTGTGGATAGGGCCCTCAGCCAGCTTCGCCCGGAGGACCCAGAGGGT	603	
Db	814	CATGCCGCCATCAGCGGGTAGCGCCGCTCATCTCCGCCACGCTCCAGCGAGGGCGTT	873	
QY	604	GACCTGCACCTAC	616	
Db	874	GGCGGCTTCTTC	886	

```

RESULT 14
US-09-252-991A-14750/c
; Sequence 14750, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14750
; LENGTH: 2505
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14750

```

Query Match	3.9%	Score 42.6;	DB 4;	Length 2505;
Best Local Similarity	51.3%;	Pred. No. 1.8;		
Matches 99;	Conservative 0;	Mismatches 94;	Indels 0;	Gaps 0;
QY 424	GTCGGCAACAGATGCTGTGGTGATCAAGGGCTTCACGGCAGAGACACGAGCGCAGTAGTACCAC	483		
DB 1824	GCTCAGTGGGGGCTGTCGTCGGGCTTTCGCACGGAACGGGGTTCGTTCAGCGG	1765		
QY 484	TGTGGCTGTGCTCAGGGCTCCATCTGCTCCCTGGGGCTGCCACCTTCCAGGTGGCAGCTGAGC	543		
DB 1764	CGTGCCAGCCCTCTCGCGCTCTTCACTTGCCTTTGAAGCCAGCACCGTTGGCACACC	1705		
QY 544	CCAGCTCTCTGTGATCAGGCCCTTCAGGCCAGCTTGCSCCCCCGAGCAGCCACGAGGCGT	603		
DB 1704	CATGGCGCCATCAGCGGTGAGCGCGTCTCATCTCCGCCACGCTCCAGCGCGGCGT	1645		
QY 604	GACCTGCACCTAC	616		
DB 1644	GGCGGTTCTTTC	1632		

```

RESULT 15
US-09-949-016-2188
; Sequence 2188, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2188
; LENGTH: 3397
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2188

```

Query Match	3.8%;	Score 42;	DB 4;	Length 3397;
Best local Similarity	51.7%;	Pred. No. 2.7;		
Matches 121;	Conservative	0;	Mismatches 110;	Indels 3;
				Gaps 1;
QY	297	CCGGCAGGTGGCAGCTGGTGAAGATGTCCTCTCGAGCTTGAGGTGGTGGCTCAGGCTGG	356	
DB	888	CCTGATGGTGGCGCAAGGACCAGGGGTGTTCAATGTGAGGTCACATGAGATGT	947	
QY	357	TGAGGTCACTTGCCACAAGGGAATGAGGCCATCCAGCCCGGTGGGCGGTTCGAGTGGT	416	
DB	948	TCGGGGTGTGTGGCTTGAAGAAATGGGAAGGAGCTGTGTGCCGACAGCGCATAAAGGTGC	1007	
QY	417	CTCCCAGGTCGGGCAACAGATGCTGTGTGATCAAGGCTTCACGGCAGAGACACCAAGGCGA	476	
DB	1008	CCATCTCGGGGGGTCCAAACTGACCATTTGACGAGTCAACCTGCCACAGAGCTGA	1067	
QY	477	GTACCACTGTGGCCTGGCTCAGGGCTCCACTGTGCCCTCGGGGTGCCACCTTCCA	530	
DB	1068	CTACAGCTTTG---TGCCCGAGGGCTTCGCTGTCAACCTGTTCAGCAAGCTCCA	1118	

Search completed: March 21, 2005, 14:39:32
Job time : 207.228 secs

Qy	121	AGCGCCAGCCGACGCTGCCCGGGAGGTGCTGGCTCGCTCGACGAGAGGCGCAGCTGC	180
Db	2120	AGCGCCAGCCGACGCTGCCCGGGAGGTGCTGGCTCGCTCGACGAGAGGCGCAGCTGC	2179
	181	TGGCTGAGCTGTCAGATCAGGCTGCGGCTGTGACGTGGCTCAAGATGGTCGCACTGT	240
Db	2180	TGGCTGAGCTGTCAGATCAGGCTGCGGCTGTGA	2239
Qy	241	CCCAAGGCCCAAGATGAGGTGAGGCATCGCGCGGCGCGGGTGTCTCTTGTGCAG	300
Db	2240	CCCAAGGCCCAAGATGAGGTGAGGCATCGCGCGGCGCGGGTGTCTCTTGTGCAG	2299
Qy	301	ATGTGGCCCGGAGCGATCGAGCTCTACAGTGGCTCAGCGCGGGGGCGCATCGCCT	360
Db	2300	ATGTGGCCCGGAGCGATCGAGCTCTACAGTGGCTCAGCGCGGGGGCGCATCGCCT	2359
Qy	361	ACCAAGCTCTCCGTGCAAGGCCCTCGCGCGCTTTCTGCAACAAGACATGGCGGGCAGCTGTG	420
Db	2360	ACCAAGCTCTCCGTGCAAGGCCCTCGCGCGCTTTCTGCAACAAGACATGGCGGGCAGCTGTG	2419
Qy	421	TGGATGCGGTGGCTGGGGCCCGGGCGAGTTTGAGTGTGAGACTTCGAGGCCACAGTCC	480
Db	2420	TGGATGCGGTGGCTGGGGCCCGGGCGAGTTTGAGTGTGAGACTTCGAGGCCACAGTCC	2479
Qy	481	ACGTGCATCTGTACAAGATGGCATGAGCTGGGGCCACTCCGGTGTAGCGCTTCTTTCAGG	540
Db	2480	ACGTGCATCTGTACAAGATGGCATGAGCTGGGGCCACTCCGGTGTAGCGCTTCTTTCAGG	2539
Qy	541	AGGATGTGGGACCGGGGACCGGCTGGTGGAGCCACAGTACACAGGAGGATGAAGGCA	600
Db	2540	AGGATGTGGGACCGGGGACCGGCTGGTGGAGCCACAGTACACAGGAGGATGAAGGCA	2599
Qy	601	CCCTACTCTGCGCGGTGGGCGAGACTCTGTGGAAGTCTCCGGCTCCGCGTCTCTGAGGCCA	660
Db	2600	CCCTACTCTGCGCGGTGGGCGAGACTCTGTGGAAGTCTCCGGCTCCGCGTCTCTGAGGCCA	2659
Qy	661	AGGTGGTGTGTTGCTTAAGAGCAGCTGSCACGACGGAAGCTGCAGGCGAGGCGAGGACCA	720
Db	2660	AGGTGGTGTGTTGCTTAAGAGCAGCTGSCACGACGGAAGCTGCAGGCGAGGCGAGGACCA	2719
Qy	721	GTGCCACACTGAGCTGCGAGGTGCCACAGCCACAGCGGAGGTACGTGTCAAGGATG	780
Db	2720	GTGCCACACTGAGCTGCGAGGTGCCACAGCCACAGCGGAGGTACGTGTCAAGGATG	2779
Qy	781	GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGGCCACAGGCTGCACGCGCAGGC	840
Db	2780	GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGGCCACAGGCTGCACGCGCAGGC	2839
Qy	841	TGGTTGTGCAGCAGGCGAGGCTGCCCGGGAGTATAGCTGCGAGGCTTGGGGGCC	900
Db	2840	TGGTTGTGCAGCAGGCGAGGCTGCCCGGGAGTATAGCTGCGAGGCTTGGGGGCC	2899
Qy	901	AGGGGCTCTCTTCCATCTGGATGTCAAGAGGCCCAAGGTGGTGTGTTGCGCAAGGACGAGG	960
Db	2900	AGGGGCTCTCTTCCATCTGGATGTCAAGAGGCCCAAGGTGGTGTGTTGCGCAAGGACGAGG	2959
Qy	961	TGGCACAACAGTGAAGTGCAGGCTGAGCGAGGGGCGCAATGCC	1001
Db	2960	TGGCACAACAGTGAAGTGCAGGCTGAGCGAGGGGCGCAATGCC	3000

RESULT 2

US-10-077-130-4

US-10-077-130-4
: Sequence 4. Application US/10077130

; sequence 4, Application US/1007
; Publication No. US20020168742A1

; Publication No. US200

APPLICANT: Kapeller-Libermann, Rosana

APPLICANT: Acton, Susan L.

APPLICANT: ACTON, SUSAN L.; TITLE OF INVENTION: 59079 and 12599, protein kinase family

1. TITLE OF INVENTION: 50075 and 123333, PROTEIN K

```

; TITLE OF INVENTION: MEMBERS AND USE
; FILE REFERENCE: MPI2001-047P1RCP1 (M)

```

; CURRENT APPLICATION NUMBER: US/10/077,130

; CURRENT FILING DATE: 2002-02-15

```
Qy 781 GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAGGCTGCACGGCGCAGGC 840
Db 2851 GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAGGCTGCACGGCGCAGGC 2910
Qy 841 TGGTTGTGACAGCAGCGAGCCAGCGGATGCCGGGGAGTATAGCTGCGAGGCTGGGGGCC 900
Db 2911 TGGTTGTGACAGCAGCGAGCCAGCGGATGCCGGGGAGTATAGCTGCGAGGCTGGGGGCC 2970
Qy 901 AGCGGCTCTCTTCATCTGATGTCAAGAGCCCAAGGTGGTGTTCGCCAAGACACAGG 960
Db 2971 AGCGGCTCTCTTCATCTGATGTCAAGAGCCCAAGGTGGTGTTCGCCAAGACACAGG 3030
Qy 961 TGGCACACAGTGAGTGTGAGCTGAGGCGAGCGGCGCAATGCC 1001
Db 3031 TGGCACACAGTGAGTGTGAGGCTGAGGCGAGCGGCGCAATGCC 3071

RESULT 3
US-10-723-860-5701
; Sequence 5701, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5701
; LENGTH: 20489
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (565)..(584)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2938)..(2955)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3329)..(3364)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4556)..(4586)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4721)..(4735)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4997)..(5011)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5108)..(5138)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6237)..(6300)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7520)..(7566)
```

```
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12014)..(12037)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12123)..(12146)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13258)..(13277)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13918)..(13942)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15219)..(15248)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15262)..(15280)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15309)..(15381)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15662)..(15682)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (17286)..(17303)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (17413)..(17430)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-5701

Query Match 98.2%; Score 983; DB 18; Length 20489;
Best Local Similarity 98.2%; Pred. No. 3.8e-243;
Matches 983; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 CGCTCACCATCCGGGAGGTGCCCGCAGCTGCGAGGTCGGGGGGCCCCAGGGCTGACTGCCAACA 60
Db 2071 CGCTCACCATCCGGGAGGTGCCCGCAGCTGCGAGGTCGGGGGGCCCCAGGGCTGACTGCCAACA 2130
Qy 61 ACGGCATTGAGAGCAGCATCCGGATGGAGGTCCGGGGGGCCCCAGGGCTGACTGCCAACA 120
Db 2131 ACGGCATTGAGAGCAGCATCCGGATGGAGGTCCGGGGGGCCCCAGGGCTGACTGCCAACA 2190
Qy 121 AGCCGCCAGCCGAGCTGCCCGGAGGTGCTGGCTCGGCTGCACAGAGAGGGCGAGCTGC 180
Db 2191 AGCCGCCAGCCGAGCTGCCCGGAGGTGCTGGCTCGGCTGCACAGAGAGGGCGAGCTGC 2250
Qy 181 TGGCTGAGCTGTCAAGTACAGGCTCGGCTGTGAGTGGCTGAAAGATGGTGCACACTGT 240
Db 2251 TGGCTGAGCTGTCAAGTACAGGCTCGGCTGTGAGTGGCTGAAAGATGGTGCACACTGT 2310
Qy 241 CCCCAGGCCCAAGTATGAGGTGAGGTCAGGCATCGGCGGGGGCCCCAGGGCTGCTCTTGTGCGAG 300
Db 2311 CCCCAGGCCCAAGTATGAGGTGAGGTCAGGCATCGGCGGGGGCCCCAGGGCTGCTCTTGTGCGAG 2370
Qy 301 ATGTGGCCCGGAGCAGTACAGGCTCTACAGTGGGTGAGGTCAGCGGGGGGGCGGCTCGCT 360
Db 2371 ATGTGGCCCGGAGCAGTACAGGCTCTACAGTGGGTGAGGTCAGCGGGGGGGCGGCTCGCT 2430
Qy 361 ACCAGCTCTCCGTGCAAGGCTCGCGGCTTTTGTGCAAGGACATGCGGCGAGCTGTG 420
```

Db 2431 ACCAGCTCTCCGTGCAAGGCTTCGCGCCTTTCTGCAAGGACATGCGGGCAGCTGTG 2490
Qy 421 TGGATGCCGTGGCTGGGGCCCGCGCAGTTTGTAGTGTGAGACCTTCGGAAGCCACAGTCC 480
Db 2491 TGGATGCCGTGGCTGGGGCCCGCGCAGTTTGTAGTGTGAGACCTTCGGAAGCCACAGTCC 2550
Qy 481 AGTGTCACTGTGTACAAAGGATGSCATGGAGCTGGGCCACTCCGGTGTAGCGCTTCTTTCAGG 540
Db 2551 ACGTGTCACTGTGTACAAAGGATGSCATGGAGCTGGGCCACTCCGGTGTAGCGCTTCTTTCAGG 2610
Qy 541 AGGATGTGGGACCGCGCCACCGCTGTGTGGAGCCACAGTACACAGGAGGATGAGGCA 600
Db 2611 AGGATGTGGGACCGCGCCACCGCTGTGTGGAGCCACAGTACACAGGAGGATGAGGCA 2670
Qy 601 CCTACTCTCCGTGGCGGCGAGGACTCTGTGGACTTCCGGCTCCCGCTCTCTGAGCCCA 660
Db 2671 CCTACTCTCCGTGGCGGCGAGGACTCTGTGGACTTCCGGCTCCCGCTCTCTGAGCCCA 2730
Qy 661 AGGTGGTGTGTCTAAGGAGCAGCTGGGCACGACGGAAGCTGCAGGACGAGGAGGCA 720
Db 2731 AGGTGGTGTGTCTAAGGAGCAGCTGGGCACGACGGAAGCTGCAGGACGAGGAGGCA 2790
Qy 721 GTGCCACACTGAGCTGCAGGTGGCCAGGCCCAGGCCCAGAGGTGACGTGTACAAAGATG 780
Db 2791 GTGCCACACTGAGCTGCAGGTGGCCAGGCCCAGGCCCAGAGGTGACGTGTACAAAGATG 2850
Qy 781 GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGATGGAGGCCACAGGCTGCACGCGCAGGC 840
Db 2851 GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGATGGAGGCCACAGGCTGCACGCGCAGGC 2910
Qy 841 TGGTGTGCACAGGACGCGCAGCGGATGCCGGGAGTATAGCTGCGAGGCTGGGGGCC 900
Db 2911 TGGTGTGCACAGGACGCGCAGCGGATGCCGGGAGTATAGCTGCGAGGCTGGGGGCC 2970
Qy 901 AGCGCTCTCTCCATCTGATGTCAAGAGCCCAAGGTGTGTTCGCCAAGACACAGG 960
Db 2971 AGCGCTCTCTCCATCTGATGTCAAGAGCCCAAGGTGTGTTCGCCAAGACACAGG 3030
Qy 961 TGGCACACAGTGAGGTGACGCTGAGGCGAGGCGCAATGCC 1001
Db 3031 TGGCACACAGTGAGGTGACGCTGAGGCGAGGCGCAATGCC 3071

RESULT 4
US-10-988-426
; Sequence 426, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyan
; APPLICANT: Wang, Dunrui
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt FL_genes Version 2.0
; SEQ ID NO 426
; LENGTH: 2768
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72) .. (2738)
US-10-120-988-426

Query Match 83.0%; Score 830.8; DB 17; Length 2768;

Best Local Similarity 95.3%; Pred. No. 5.4e-204;
Matches 856; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
Qy 1 CGCTCACATCCGGAGAGTGCCTCCGCGAGCTGACAGGGGCGCAGCTGAAGTTCGTGGCCA 60
Db 1861 CGCTCACATCCGGAGAGTGCCTCCGCGAGCTGACAGGGGCGCAGCTGAAGTTCGTGGCCA 1920
Qy 61 ACGCCATTGAGAGCAGCATCCGGATGAGGTTCGGGCGGGCCAGGGCTGACTGCCAACA 120
Db 1921 ACGCCATTGAGAGCAGCATCCGGATGAGGTTCGGGCGGGCCAGGGCTGACTGCCAACA 1980
Qy 121 AGCCGACGCGCAGCTGCCCGGAGGTGCTGGCTCGGCTGCAACAGGAGCGCAGCTGC 180
Db 1981 AGCCGACGCGCAGCTGCCCGGAGGTGCTGGCTCGGCTGCAACAGGAGCGCAGCTGC 2040
Qy 181 TGGCTGAGCTGTGATGATCAGCTGCGGCTGTGAGCTGGCTGGAAGGATGCTGCGACACTGT 240
Db 2041 TGGCTGAGCTGTGATGATCAGCTGCGGCTGTGAGCTGGCTGGAAGGATGCTGCGACACTGT 2100
Qy 241 CCCCAGGCCCCAAAGTATGAGGTGCAGGCATCGGCGGGCGGGGCTCTCTTGTGCGAG 300
Db 2101 CCCCAGGCCCCAAAGTATGAGGTGCAGGCATCGGCGGGCGGGGCTCTCTTGTGCGAG 2160
Qy 301 ATGTGGCCCGGAGCGATGACAGGCTCTTACGAGTGCCTCAGCCGCGGGGCGGCATCGCCT 360
Db 2161 ATGTGGCCCGGAGCGATGACAGGCTCTTACGAGTGCCTCAGCCGCGGGGCGGCATCGCCT 2220
Qy 361 ACCAGCTCTCCGTGACAGGCTTCGCGGCTTTCTGCAACAGGACATGCGGGGAGCTGTG 420
Db 2221 ACCAGCTCTCCGTGACAGGCTTCGCGGCTTTCTGCAACAGGACATGCGGGGAGCTGTG 2280
Qy 421 TGGATGCGGTGGCTGGGGGCGCGCAGTTTTCAGTGTGAGACTCCGAAAGCCACACTCC 480
Db 2281 TGGATGCGGTGGCTGGGGGCGCGCAGTTTTCAGTGTGAGACTCCGAAAGCCACACTCC 2340
Qy 481 ACFTGCACTGGTAAAGGATGGCATGGAGTGGGCCACTCCGGTGAAGCGCTTCTTTCAGG 540
Db 2341 ACFTGCACTGGTAAAGGATGGCATGGAGTGGGCCACTCCGGTGAAGCGCTTCTTTCAGG 2400
Qy 541 AGGATGTGGGAGCGCGCACCGGCTGTGTGACGACACAGTCCACAGGACAGGATGAGGCA 600
Db 2401 AGGATGTGGGAGCGCGCACCGGCTGTGTGACGACACAGTCCACAGGACAGGATGAGGCA 2460
Qy 601 CCTACTCTCCGCGGTGGGCGAGGACTCTGTGACTTCCGGCTCCGGCTCTCTGAGGCCA 660
Db 2461 CCTACTCTCCGCGGTGGGCGAGGACTCTGTGACTTCCGGCTCCGGCTCTCTGAGGCCA 2520
Qy 661 AGGTGTGTGTCTAAGGAGCAGCTGGSCACGCAAGGAGCTGCAGGACAGGAGGAGGCA 720
Db 2521 AGGCGGTGTGTTCGCAAGGAGCAGCGCGCTGACAGGAGGTGACAGGCTGAGGTGGGGGCA 2580
Qy 721 GTGCCACACTGAGCTGCGAGGTGGCCCGCCAGCGGAGGTGACGTGTGTACAAAGATG 780
Db 2581 GCGCCACGCTGAGCTGTGAGGTGGCCCGCCAGGACACAGATGGAGGTGACGTGTGTACAAAGATG 2640
Qy 781 GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGATGGAGGCCACAGGCTGCACGCGCAGGC 840
Db 2641 GGAAGAAGCTGAGCTCCAGCTCGAAAGTGTGATGGAGGTGAGGTGTGTGCGGAGGC 2700
Qy 841 TGGTGTGACAGCAGGCGGAGTGGCGGAGTGGCGGGAGTATAGCTGCGAGGCTGGGGG 898
Db 2701 TGGTGTGACAGCAGGCGGAGTGGCGGAGTGGCGGAGTGGCGGGAGTATAGCTGCGAGGCTGGGGG 2758

RESULT 5
US-10-027-632-272054
; Sequence 272054, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129

```
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272054
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-272054

Query Match      19.1%; Score 191; DB 13; Length 578;
Best Local Similarity 89.2%; Pred. No. 1.8e-39;
Matches 206; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 650 CTCAGAGCCCAAGTGGTGTGTTTAAAGGAGCAGCTGGCAGCAGGAAAGTGCAGGCGAGA 709
DB 347 CCCAGAGCCCAAGTGGTGTGTTTAAAGGAGCAGCTGGCAGCAGGAAAGTGCAGGCGAGA 406

QY 710 GGCAGAGCCCAAGTGGTGTGTTTAAAGGAGCAGCTGGCAGCAGGAAAGTGCAGGCGAGA 769
DB 407 GGCAGAGCCCAAGTGGTGTGTTTAAAGGAGCAGCTGGCAGCAGGAAAGTGCAGGCGAGA 466

QY 770 GTACAAGGATGGGAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCGCACAGGCTG 829
DB 467 GTACAAGGATGGGAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCGCGTGGGCTG 526

QY 830 CACGCGCAGGCTGTTGTGTCAGCAGGAGCGAGGCGGATCCCGGGAGTA 880
DB 527 CACGCGCAGGCTGTTGTGTCAGCAGGAGCGGCGGATCCCGGGAGTA 577

RESULT 6
US-10-027-632-272054
; Sequence 272054, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272054
```

```
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-272054

Query Match      19.1%; Score 191; DB 17; Length 578;
Best Local Similarity 89.2%; Pred. No. 1.8e-39;
Matches 206; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 650 CTCAGAGCCCAAGTGGTGTGTTTAAAGGAGCAGCTGGCAGCAGGAAAGTGCAGGCGAGA 709
DB 347 CCCAGAGCCCAAGTGGTGTGTTTAAAGGAGCAGCTGGCAGCAGGAAAGTGCAGGCGAGA 406

QY 710 GGCAGAGCCCAAGTGGTGTGTTTAAAGGAGCAGCTGGCAGCAGGAAAGTGCAGGCGAGA 769
DB 407 GGCAGAGCCCAAGTGGTGTGTTTAAAGGAGCAGCTGGCAGCAGGAAAGTGCAGGCGAGA 466

QY 770 GTACAAGGATGGGAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCGCACAGGCTG 829
DB 467 GTACAAGGATGGGAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCGCGTGGGCTG 526

QY 830 CACGCGCAGGCTGTTGTGTCAGCAGGAGCGAGGCGGATCCCGGGAGTA 880
DB 527 CACGCGCAGGCTGTTGTGTCAGCAGGAGCGGCGGATCCCGGGAGTA 577

RESULT 7
US-09-726-643-22
; Sequence 22, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-643-22

Query Match      8.8%; Score 88.4; DB 9; Length 1645;
Best Local Similarity 46.5%; Pred. No. 4.3e-13;
Matches 358; Conservative 0; Mismatches 406; Indels 6; Gaps 2;

QY 165 GAGGAGCGCAGCTGCTGGCTGAGCTGCAGATCAGGCTGGCGTGTGACCTGGCTGAAG 224
DB 581 GAGCGGGTGGTGGCTGACTTGTGAGCTCTCAGGGTGGACTTCCCGGCAACCTGGTACAAG 640

QY 225 GATGCTCGCACATGTGTCAGGCGCCCAAGTATGAGTGCAGGCAATCGCGCGGCGCGG 284
DB 641 GATGGCGAAGAGTGGAGGAGAGCGAGTGTGCTGGTGGTGAAGATGATGGCGCAACAC 700

QY 285 GTGCTCTCTTGTGCGAGATGTCGCGCGGAGCAGTGCAGGCTCTACGAGTGGCTCAGCGC 344
DB 701 CGTCTGATCTCTGCGCTGAGGCGCAAGTCCAGACAGTGGCGAGTTTGAATGCGAGGAC 760

QY 345 GGGGCGCGCATCGCTTACAGCTCTCGCTGCAAGCGCTCGCGCGCTTTCTGCACAAAGAC 404
DB 761 GGGGTCTCGGCTTCTTGGCGCTCACTGTCCAAATCTCTCCGTCGACATCGTGAACCCC 820

QY 405 ATGGCGGGCAGCTGTGTGATGTCGCTGGCTGGCGCGCGCGGCGAGTTTGAATGAGTGA 464
DB 821 CGAGAACATGTGTTGTCGTGCATGCCATAACTTCCGAGTGTGTGTCATGCTGCGCTGTG 880

QY 465 TCCGAAGCCACAGCTCCAGCTGCTACTGGTACAAGGATGCGATGGAGCTGGGCCACTCCG 524
```

```
Db 881 GACCGAGAGGACGCCCTCTGTGGTTGTTACAAAGACGGGCGAGAGGTGGAGGAGA---GT 937
Qy 525 GAGCGCTTCTTTCAGGAGGATGTGGGACGGGACACCGGCTGTGGGACGACAGTCAACC 584
Db 938 GACTTGTGTGCTTGGAGAAATGAGGGGCCCATCGCCGCTGTGGTCTGCCCGCCACCCAT 997
Qy 585 AGGAGGATGAAGGACCTACTCTCTGCGCGTGGGCGAGGACTCTGTGGACTTCCGGGTC 644
Db 998 CCCTCAGACGGGGCGAGTTTCAGTGTGCTGTGAGATGAGTGTGCCCTACTTCACTGTC 1057
Qy 645 CGGTCTCTAGGCCAAGGTGGTTTCTTAAGAGACAGTGTGGCAGCAGGAAAGCTGAG 704
Db 1058 ACCATCAGACGCTCTCTCTGTGGATCG---TGTATCCAGCGGCAAGGTGTATGTGGCA 1114
Qy 705 GCAGAGGAGGAGCCAGTGCACACTGAGTGTGAGGTGGCCAGGCGCCAGCAGGAGGTG 764
Db 1115 GCGTGTGCGCTGGAGCGTGTGGTGTGACCTGTGAGCTATCCGGGCTTGGGCGAGAGGTG 1174
Qy 765 ACGTGTACAAAGATGGGAAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACA 824
Db 1175 CGCTGGACCAAGGATGGAGAGGAGTGTGGAGAGCCCGCGCTGCTCTGCAAGGAA 1234
Qy 825 GGCTGCACGCGCAGGCTGTGTGAGCAGGAGGCGGCGAGTGGCGGAGATATAGC 884
Db 1235 GACACTGTCCCGCCTGTGTCTGCCGCTGTCAGGCTTCAGAGGACTCCGGCGAGTACTTG 1294
Qy 885 TGCAGGCTGGGGCCAGGCGCTCTCTCTTCATCTGGATGTCAAGAGCC 934
Db 1295 TGTGAATTGACGATGAGTCGGCCTCTTCACTGTACCCGTACAGAGTC 1344
```

RESULT 8

```
US-10-042-141-22
; Sequence 22, Application US/10042141
; Publication No. US20020183503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/10/042,141
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-141-22
```

Query Match 8.8%; Score 88.4; DB 13; Length 1645;
Best Local Similarity 46.5%; Pred. No. 4.3e-13;
Matches 358; Conservative 0; Mismatches 406; Indels 6; Gaps 2;

```
Qy 165 GAGGAGCGCAGCTGTGGCTGTGAGCTGTCAAGTACAGGCTGCGGCTGTGAGCTGCTGAAG 224
Db 581 GAGCGGTGTGTGCTGACTTGTGAGCTCTCAAGGGTGGACTTCCCGGCAACCTGGTACAAG 640
Qy 225 GATGGTCGCACACTGTCCCGAGGCGCCCAAGTATAGGTGAGGATCGGCGGCGCGG 284
Db 641 GATGGCAGAGGTGGAGGAGAGCGAGTGTGCTGTGTGAAGATGGGATGGCGCAACAC 700
Qy 285 GTGCTCTCTGTGCGAGATGTGGCCCGGACGATGCAGGCGCTCTACGAGTGGTCTCAGCGC 344
Db 701 CGTCTGATCTCTGCTGAGGCCCAAGTCCAGACAGTGTGGCGAGTTTGTAGTGCAGGACAG 760
Qy 345 GGGGGCCGCATCGCCTTACGAGCTCTCCGTGCAAGGCGCTCGCGCGCTTCTTCTGCAAGGAC 404
```

```
Db 761 GGGGTCTCGGCTTCTTTCGGGCTCACTGTCCAAGATCCTCCCGTGCAATCGTGGACCCC 820
Qy 405 ATGCGCGGCGCAGCTGTGTGGATTCGGTGTGGGCGCCGCGGCGAGTTTGTAGTGTGAGACC 464
Db 821 CGAGAAATGTGTTCGTGCTATGCTTAACATTCGAGTGTGTATCTGTGGCTGTGAGGTG 880
Qy 465 TCCGAAGCCCAACGCTCCACGTCACCTGGTACAAAGATGGCATGGAGCTGGGCGCACTCCCGT 524
Db 881 GACCGAGAGGACGCCCTCTGTGGTTGGTACAAAGACGGGCGAGGAGGTGGAGGAGA--GT 937
Qy 525 GAGCGCTTCTTTCAGAGGAGATGTGGGACCGGGCACCGGCTGTGGTGGGACGACAGTCAACC 584
Db 938 GACTTCTGTGTGTGGAGAAATGAGGGGCCCATCGCGGCTGTGTGCTGCTGCCCGCCACCCAT 997
Qy 585 AGGAGGATGAAGGACCTACTCTCTCGCGGTGGGCGAGGACTCTGTGGACTTCCGGGTC 644
Db 998 CCCTCAGACGGGGCGAGTTTCAGTGTGCTGTGAGATGAGTGTGCTTCACTTCTTCACTGTC 1057
Qy 645 CGGTCTCTGAGCCCAAGGTGGTGTGTTGCTAAGAGCAGCTGGGCAAGGAGTGCAG 704
Db 1058 ACCATCAGACGCTCTCTCTGTGGATCG---TGTATCCAGCGGCAAGGTGTATGTGGCA 1114
Qy 705 GCAGAGCAGAGGACCACTGCCCACACTGAGCTGTGAGTGTGGGCGGCGGCGGCGAGGAGTG 764
Db 1115 GCGTGTGCGCTTGGAGCGTGTGTGTGCTGACCTGTGAGCTATGCGCGGCTTGGGCGAGAGTG 1174
Qy 765 ACGTGTGTACAAAGATGGGAAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACA 824
Db 1175 CGCTGGACCAAGGATGGAGAGGAGTGTGGAGAGGCGCCGCGCTGCTCTGCAAGGAA 1234
Qy 825 GGCTGCACGCGCGCAGGCTGTGTGTCAGCAGGCGGCGGCGGATGCGGGGAGATATAGC 884
Db 1235 GACACTGTCCCGCCTGTGTGCTGCCGCTGTCCAGCTTCGAGGACTCCGGCGAGTACTTG 1294
Qy 885 TGCAGGCTGGGGCCAGGCGGCTCTCTTCCATCTGGATGTCAAGAGCC 934
Db 1295 TGTGAATTGACGATGAGTCGGCCTCTTCACTGTCAACCGTTCACAGAGTC 1344
```

RESULT 9

```
US-10-919-272-22
; Sequence 22, Application US/10919272
; Publication No. US20050010042A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/10/919,272
; CURRENT FILING DATE: 2004-08-17
; PRIOR APPLICATION NUMBER: US/09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-919-272-22
```

Query Match 8.8%; Score 88.4; DB 19; Length 1645;
Best Local Similarity 46.5%; Pred. No. 4.3e-13;
Matches 358; Conservative 0; Mismatches 406; Indels 6; Gaps 2;

```
Qy 165 GAGGAGCGCAGCTGTGGCTGTGAGCTGTGAGTCAAGTACAGGCTGCGGCTGTGAGCTGCTGAAG 224
Db 581 GAGCGGTGTGTGCTGACTTGTGAGCTCTCAAGGGTGGACTTCCCGGCAACCTGGTACAAG 640
Qy 225 GATGGTCGCACACTGTGTCCCGAGGCGCCCAAGTATAGGTGAGGATCGGCGGCGGCGG 284
```

Db 641 GATGGGAGAGGTGGAGGAGCGAGTTGCTGTGTGTGATGATGGATGGCGCAACAC 700
Qy 285 GTGCTCCTTGTGCGAGATGTGGCCCGGACGATCGAGCCCTCTACGATGGTGTGAGCGGC 344
Db 701 CGTCTGATCTCTGCTGAGGCGCAAGTCCAGGACAGTGGCGAGTTTGAAGTGCGAGACAGAA 760
Qy 345 GGGGGCGCATCGCTACCAAGCTCTCGGTGCAAGGCTCCGCGCTTCTGCAACAGGAC 404
Db 761 GGGGCTCTGGGCTTCTTGGGCTCACTGTCCAAGATCTCCCGTGCAATCGTGGACCCC 820
Qy 405 ATGCGGGCGCAGCTGTGTGGATGCGCTGGGGCGCGCGGCGAGTTTGAAGTGAGAGTGC 464
Db 821 CGAGAACATGTGTGCGTGCATGCCATAACTTCCAGTGTGTATGCTGGCTGTGAGGTG 880
Qy 465 TCCGAAGCCACGTCACGTGCACTGTGTACAGGATGGCATGGAGCTGGGCGCATCTCGGT 524
Db 881 GACCGAGAGGACGCCCTGCTGCGTTGGTACAAAGACGGGCGAGGAGGTGGAGGAGA--GT 937
Qy 525 GAGCGCTTCTGCAAGGAGGATGTGGGACGGGACCGGCTGGTGGGACGACAGTCACC 584
Db 938 GACTTCTGTGTGTGGAGAAATGAGGGGCGCCCATCGCGCTGGTGTGCGCGCCACCCAT 997
Qy 585 AGGCGAGTGAAGGCACCTACTCTCGCGGTGGCGGAGGACTCTGTGGACTTCCGGCTC 644
Db 998 CCCTCAGACGGGGCGAGTTTCACTGCTGCTGGAGATGATGTGCTACTTCACTGTC 1057
Qy 645 CGGCTCTGTAGCCCAAGGTGTGTTGCTAAGAGCAGCTGGCAGCGAGAACTGCGAG 704
Db 1058 ACCATCAGACAGCTCTCTCTGTGGATCG---TGTATCCAGCGCAAGGTGTATGTGCA 1114
Qy 705 GCAGAGCAGGAGCCAGTGCACACTGAGTGCAGGTGGCGGAGGCGCCAGCGAGGAGTG 764
Db 1115 GCGGTGCGCTGGAGCGTGTGTGCTGACCTGTGAGCTATGCGCGGCGCTGGGCGAGAGTG 1174
Qy 765 ACGTGGTACAGGATGGGAAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACA 824
Db 1175 CGCTGGACCAAGATGAGAGGAGGTGTGAGAGCGCCGCGCTGCTCTGCAGAGGAA 1234
Qy 825 GGCTGACGCGCAGGCTGTGTGTCAGCAGCGACGCGGATGCGGGGAGTATAGC 884
Db 1235 GACACTGTCCGCGCGCTGTGCTGCGCTGTCAGCTCGAGGACTCGGCGAGTACTTG 1294
Qy 885 TCGAGGCTGGGGCGCCAGCGCTCTCTTCCATCTGATGTCAAGAGCC 934
Db 1295 TGTGAATTGACGATGAGTGGCGCTCTCTTCACTGTCAACCGTCCACAGATC 1344

RESULT 10

US-10-264-049-850
; Sequence 850, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birex et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 850
; LENGTH: 2170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-049-850

Query Match 8.8%; Score 88.4; DB 17; Length 2170;
Best Local Similarity 46.5%; Pred. No. 4,3e-13;
Matches 358; Conservative 0; Mismatches 406; Indels 6; Gaps 2;

Qy 165 GAGGAGCGCAGCTGCTGGCTGAGCTGTAGATCAGGCTCGGCTGTGACCTGTGCTGTAAG 224
Db 1072 GAGCGGTGTGTGCTGACTTGTGAGCTCTCAAGGGTGGACTTCCCGGCAACCTTGGTACAG 1131
Qy 225 GATGGTCCGACACACTGTCTCCAGGCCCCAAGTATAGGTTGAGGTCAGGCAATCGCGCGCGCGG 284
Db 1132 GATGGGCAAGAGTGGAGGAGGAGTGTCTGTGTGTGAAGATGGATGGGCGCAACAC 1191
Qy 285 GTGCTCCTTGTGCGAGATGTGGCCCGGACGATCGAGCCTCTACGAGTGGTGTGAGCGCG 344
Db 1192 CGTCTGATCTCTGCTGAGGCGCAAGTCCAGGACAGTGGCGAGTTTGAAGTGCGAGCAGAA 1251
Qy 345 GGGGGCGCATCGCTACCAAGCTCTCGGTGCAAGGCTCCGCGCTTCTGCAACAGGAC 404
Db 1252 GGGGTCTCGGCTTCTTCCGGCTCACTGTCCAAGATCTTCCCGTGCAATCGTGGACCCC 1311
Qy 405 ATGCGGGCGCAGCTGTGTGGATGCGCTGGGGCGCGCGGCGAGTTTGAAGTGTGAGAGC 464
Db 1312 CGAGAACATGTGTGCTGTCATGCCATAACTTCCAGTGTGTATGCTGGCCTGTGAGGTG 1371
Qy 465 TCCGAAGCCACGTCACGTGCACTGTGTACAGGATGGCATGGAGCTGGGCGCATCTCGGT 524
Db 1372 GACCGAGAGGACGCCCTGCTGCGTTGGTACAAAGACGGGCGAGGAGTGGAGGAGA--GT 1428
Qy 525 GAGCGCTTCTGCAAGGAGGATGTGGGACGGGACCGGCTGGTGGGACGACAGTCACC 584
Db 1429 GACTTCTGTGTGTGGAGAAATGAGGGGCGCCCATCGCGCTGGTGTGCTCCCGCCACCCAG 1488
Qy 585 AGGCGAGTGAAGGCACCTACTCTCTCGCGGTGGCGGAGGACTCTGTGGACTTCCGGCTC 644
Db 1489 CCCTCAGACGGGGCGAGTTTCACTGCGTGGAGATGATGTGCTACTTCACTGTC 1548
Qy 645 CGGCTCTGTAGCCCAAGGTGTGTTGCTAAGAGCAGCTGGCAGCGAGGAGTGGAG 704
Db 1549 ACCATCAGACAGCTCTCTCTGTGGATCG---TGTATCCAGCGCAAGGTGTATGGCA 1605
Qy 705 GCAGAGCAGGAGCCAGTGCACACTGAGTGCAGGTGGCGGAGGAGTGGCGCCAGCGAGTG 764
Db 1606 GCGGTGCGCTGGAGCGTGTGTGCTGACCTGTGAGCTATGCGCGCTTGGCGAGAGTG 1665
Qy 765 ACGTGGTACAGGATGGGAAGCTGAGCTCCAGCTCGAAAGTGTGTGATGGAGGCCACA 824
Db 1666 CGCTGGACCAAGATGAGAGGAGGTGTGGAGAGCGCCGCGCTGCTCTGCAGAGGAA 1725
Qy 825 GGCTGACGCGCAGGCTGTGTGTCAGCAGCGACGCGGATGCGGGGAGTATAGC 884
Db 1726 GACACTGTCCGCGCGCTGTGCTGCGCTGTCCAGCTCGAGGACTCGGCGAGTACTTG 1785

RESULT 11

US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1946

Query Match 8.6%; Score 86.2; DB 17; Length 2534;

[illegible]

RESULT 12

US-10-476-397-15	
; Sequence 15, Application US/10476397	
; Publication No. US20040115687A1	
GENERAL INFORMATION:	
APPLICANT: YUE, Henry	
APPLICANT: LEE, Ernestine A.	
APPLICANT: DUGGAN, Brendan M.	
APPLICANT: THANGAVELU, Kavitha	
APPLICANT: HONCHELL, Cynthia D.	
APPLICANT: DING, Li	
APPLICANT: JACKSON, Jennifer L.	
APPLICANT: BAUGHN, Mariah R.	

Db 493 GCTTTGAGTGGCTCTGTCTCCACAGAGAGTGCCAGCGACCCGGGCCATGTGGACAGTCCGGTG 552
Qy 781 GGAAGAGAGTGTAGCTTCAGCTCGAAGCTGTGCATGGAGGCCACACAGGCTGCACGGCAGGC 840
Db 553 GGAAGACAGTGGGAGCTTCAGCGCGCTTCAGGCCACACGTGAGGGCCGAAATACATCC 612
Qy 841 TGGTTGTGACAGCAGGAGCCAGCGGATGCGGGGAGTATAGTGCAGGCTGGGGGCC 900
Db 613 TGGTGGTCCGGAGGCTGCACCAAGTGTGCGGGAGGTGGTCTTCTGTGCGGGGCC 672
Qy 901 AGCGGCTCTCTTCCATCTGTGATGTCAAAGAGCCCAAGGTGGTGTTCGCAAGACAGG 960
Db 673 TCACCTCAAGGCTCACTCACTTGTCAAGAGAGCGCGGCCCATCATCAAGCCCTGG 732
Qy 961 TGGCACACAGTGTAGTGCAGGCTCAGGCGAGGG 993
Db 733 AGACCAGTGGTGGCGCCAGGAGGACGTGG 765

RESULT 14

US-10-093-463-71

; Sequence 71, Application US/10093463

; Publication No. US20030208039A1

; GENERAL INFORMATION:

; APPLICANT: Padigar, Muralidhara

; APPLICANT: Shenoy, Suresh

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Gusev, Vladimir

; APPLICANT: Pochart, Pascal

; APPLICANT: Zhong, Mei

; APPLICANT: Rastelli, Luca

; APPLICANT: Mezes, Peter

; APPLICANT: Smithson, Glennda

; APPLICANT: Guo, Xiaojia

; APPLICANT: Gerlach, Valerie

; APPLICANT: Casman, Stacie

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Tchernev, Velizar

; APPLICANT: Gangolli, Esha

; APPLICANT: Vernet, Corine

; APPLICANT: Pena, Carol

; APPLICANT: Burgess, Catherine

; APPLICANT: Liu, Xiaohong

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gorman, Linda

; APPLICANT: Spaderna, Steven

; APPLICANT: Voss, Edward

; APPLICANT: Malyankar, Uriel

; APPLICANT: Anderson, David

; APPLICANT: Patturajan, Meera

; APPLICANT: Miller, Charles

; APPLICANT: Taupier, Raymond J. Jr.

; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypepti

; TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.

; FILE REFERENCE: 21402-290A (Cura 590AT)

; CURRENT APPLICATION NUMBER: US/10/093,463

; PRIOR FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 60/283,675

; PRIOR FILING DATE: 2001-04-14

; PRIOR APPLICATION NUMBER: 60/338,092

; PRIOR FILING DATE: 2001-12-03

; PRIOR APPLICATION NUMBER: 60/274,281

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: 60/274,101

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: 60/325,681

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/304,354

; PRIOR FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/279,995

; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 14109
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(14088)
US-10-093-463-71

Query Match 7.9%; Score 78.6; DB 17; Length 14109;

Best Local Similarity 48.5%; Pred. No. 1.4e-10;

Matches 278; Conservative 0; Mismatches 289; Indels 6; Gaps 2;

Qy 421 TGGATGCGGTGGGGCCCGCGCAGTTTGAAGTGTGAGACTCCGAGCCACGTCC 480
Db 199 TGTCCGACAGGAGCGCGCAGCTTGGCCCTGCGAGTGTGAAGTCTCTGACCCGAGGCC 258
Qy 481 ACGTGACTGTGTACAGGATGGCATGGAGCTGGGCCACTCGGTGAGCGCTTCTTGAGG 540
Db 259 ATGTGTGTGGCGCAAGATGGCGTGCAGTGGGCCCCAGTGACAAAGTATGACTTCTGTC 318
Qy 541 AGGATGTGGGACCGCGCACCGCTGTGTGCAGCCACAGTCACAGGACGAGTCAAGGCA 600
Db 319 ACAGCGCGGCACCGGG---GGCTGTGTGTGATGACGTGAGCCCTGAAGACCGCGGCC 375
Qy 601 CCTACTCTCGCGGTGGGCGAGGACTCTGTGAGACTTCCGGCTCCGGCTCTCTGAGGCCA 660
Db 376 TGTACACCTGCCACGTGGGCTCCGAGGAGACCCGGGCCCGGGTCCGGGTGACGATCTGC 435
Qy 661 AGGTGTTGTTTGAAGGACAGTGGCAGCGAGGAAGCTGCAGGCGAGGAGCAGGAGCCCA 720
Db 436 ACGTGGGCATCACCAAGAGGCTGAAAGACA---ATGGAGGTGCTGGAAGGGGAAAGCTGCA 492
Qy 721 GTGCCACACTGAGTGGGAGGTGGCCCGCCAGCGAGGTCACGTGGTGTACAGGATG 780
Db 493 GCTTTGAGTGGTCTGTCTCCACGAGAGTGCACGACCCGGCCATGTGGACAGTCTGGTG 552
Qy 781 GGAAGAGCTGAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAGGTGACGCGCAGGC 840
Db 553 GGAAGACAGTGGGACGCTCCAGCGGCTTCCAGGCCACACGTCAGGCGCGAAATATCATCC 612
Qy 841 TGGTTGTGACAGGAGGCGCAGCGGATCCGGGGAGTATAGTGTGAGGCTGGGGGCC 900
Db 613 TGGTGGTCCGGAGGCTGCACCAAGTGTATCCGGGGAGGTGGTCTTCTGTGCGGGGCC 672
Qy 901 AGCGGCTCTCTTCCATCTGGATGTCAAAGAGCCCAAGGTGGTGTTCGCAAGACAGG 960
Db 673 TCACCTCAAGGCTCACTCATTTGTCAAGAGAGGCGCGGCCCATCATCAAGCCCTGG 732

Qy 961 TGGCACAGTGAGGTGAGCTGAGCGAGGG 993
Db 733 AAGACCAGTGGTGGCGCCAGGGGAGACGTGG 765

RESULT 15

US-09-726-643-44
; Sequence 44, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 2254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-643-44

Query Match 6.8%; Score 67.8; DB 9; Length 2254;
Best Local Similarity 49.7%; Pred. No. 8.5e-08;
Matches 256; Conservative 0; Mismatches 252; Indels 7; Gaps 3;
Qy 420 GTGATGCCGTGGCTGGGGGCGCGCAGCTTTGAGTGTGAGACCTCGAAGCCACGTC 479
Db 1331 GTGATGCCAACTTCCGAGTGTGTCTATGCTGGCCTGTGAGGTGGACCGAGAGGACGCC 1390
Qy 480 CACGTGCACTGTCACAGGATGGCATGAGCTGGGCCACTCCGGTGAGCGCTTCTTGCGAG 539
Db 1391 CCTGTGCTTGGTACAGGACGGGAGGAGGTGG---AGGAGATGACTTCTGTGTGTCTG 1447
Qy 540 GAGGATGTGGGAGCGCGCACCGGTGTGGCAGCCACAGTCACCAGGCAGGATGAAGGC 599
Db 1448 GAGAATGAGGGGCCCCCATCGCGCTGTGTGCCCGCCACCAGCCCTCAGACGGGGGC 1507
Qy 600 ACCTACTCCTCGCGGTGGGCGAGACTCTGTGAGACTTCGGGCTCCGGCTCTCTGAGCCC 659
Db 1508 GAGTTTCAGTCGCTGCTGGAGATGAGTGTGCTACTTCACTGTCAACATCACAGACGTC 1567
Qy 660 AAGTGTGTGTTTGTAAAGGACAGCTGGCACGAGGAGCTGCAGGCAGAGCAGGAGCC 719
Db 1568 TCCTCGTGGATCGT---GTATCCAGCGGCAAGGTGTATGTGGCAGCCGTGCGCCTGGAG 1624
Qy 720 AGTGCCACACTGAGCTGCGAGGTGGCCCGAGCCAGAGGAGTACGTTGGTACAAAGGAT 779
Db 1625 CGTGTGTGCTGACCTGTGAGCTATGCCGGCCCTGGGCGAGAGGTGCGCTGGACCAAGAT 1684
Qy 780 GGAAGAGAGCTGAGCTCCAGCTCGAAAGTGTGATGAGAGGCCACAGGCTGCACGCGCAGG 839
Db 1685 GGAGAGGAGGTGGTGGAGACCCCG-CGCTGCTCTCTGCAGAGGAAGACACTGTTCGCGCGC 1743
Qy 840 CTGCTTGTGACGAGGACGAGCGGATGCCGGGAGTATAGCTGCGAGGCTGGGGGC 899
Db 1744 CTGCTGCTGCGCGGTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAAATTGACGAT 1803
Qy 900 CAGCGGCTCTCTTCCATCTGGATGTCAAAGAGCC 934
Db 1804 GAGTCGGCTCTCTTCACTGTCAACCTCACAGATC 1838

Search completed: March 21, 2005, 16:07:49
Job time : 631.052 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:32:32 ; Search time 1879.9 Seconds
(without alignments)
9503.485 Million cell updates/sec

Title: US-10-077-130-6_COPY_19500_22500
Perfect score: 3001
Sequence: 1 cctacggagcagaactcggg.....ccacctgggagcaagacgga 3001

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5544816 seqs, 2976611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3001	100.0	4936	17	US-10-182-243-14
2	3001	100.0	7893	13	US-10-077-130-3
3	3001	100.0	8106	13	US-10-077-130-1
4	3001	100.0	23907	13	US-10-077-130-6
5	3001	100.0	24120	13	US-10-077-130-4
6	2999.4	99.9	5007	15	US-10-307-019-3
7	2999.4	99.9	7928	15	US-10-307-019-5
8	2997.8	99.9	5207	9	US-09-858-664A-1
9	2997.8	99.9	5207	14	US-10-274-978-1
10	2997.8	99.9	5207	14	US-10-274-978-3
11	2997.8	99.9	5207	17	US-10-697-263-1

12	2997.8	99.9	5207	17	US-10-697-263-3
13	2997.8	99.9	5207	18	US-10-921-168-1
14	2997.8	99.9	5207	18	US-10-921-168-3
15	2996.2	99.8	5454	17	US-10-415-011-44
16	2643.4	88.1	4175	15	US-10-307-019-2
17	1706.8	56.9	3225	17	US-10-311-034-46
18	223.2	7.4	7789	17	US-10-333-314-38
19	223.2	7.4	9698	17	US-10-210-130-37
20	223.2	7.4	9807	17	US-10-379-381-1
21	204	6.8	9930	17	US-10-210-130-33
22	204	6.8	10122	17	US-10-210-130-35
23	173.6	5.8	860	17	US-10-210-130-39
24	170.2	5.7	2226	18	US-10-335-053-219
25	168.6	5.6	2105	18	US-10-468-406-17
26	168.6	5.6	2132	17	US-10-262-443-118
27	168.6	5.6	2224	14	US-10-059-585-11
28	163	5.4	382	11	US-09-864-408A-8947
29	148.6	5.0	1467	17	US-10-425-114-16427
30	148.6	5.0	3164	16	US-10-210-120-25
31	148.6	5.0	4564	10	US-09-814-353-21660
32	148.6	5.0	4611	18	US-10-357-930-23396
33	148.6	5.0	4611	18	US-10-357-930-24900
34	148.6	5.0	4611	18	US-10-357-930-29283
35	148.6	5.0	5166	19	US-10-741-600-355
36	148.6	5.0	5926	9	US-09-969-708-302
37	148.6	5.0	5926	9	US-09-954-456-522
38	148.6	5.0	5926	9	US-09-880-107-3371
39	148.6	5.0	5926	10	US-09-873-367C-247
40	148.6	5.0	5926	10	US-09-873-367C-787
41	148.6	5.0	5926	17	US-10-240-425-1440
42	148.6	5.0	6202	19	US-10-741-600-363
43	148.6	5.0	7359	19	US-10-741-600-365
44	148.6	5.0	7478	19	US-10-741-600-368
45	148.6	5.0	7566	19	US-10-741-600-357

ALIGNMENTS

RESULT 1

US-10-182-243-14
; Sequence 14, Application US/10182243
; Publication No. US20040048310A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY D.
; APPLICANT: WHYTE, DAVID
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; APPLICANT: MARTINEZ, RICARDO
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE
; FILE OF INVENTION: ENZYMES
; FILE REFERENCE: 038602/1366
; CURRENT APPLICATION NUMBER: US/10/182,243
; CURRENT FILING DATE: 2003-07-07
; PRIOR APPLICATION NUMBER: PCT/US01/02337
; PRIOR FILING DATE: 2001-01-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 4936
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-182-243-14

Query Match	100.0%;	Score 3001;	DB 17;	Length 4936;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3001;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	CCTACGGAGCAGAACTCGGCCCGCAGCATACAGGACGAGACATCTTGGCCGCTGAG	60	
Db	450	CCTACGGAGCAGAACTCGGCCCGCAGCATACAGGACGAGACATCTTGGCCGCTGAG	509	
Qy	61	CCACCCCTGCTACCGGGGCTGCTGGACCACTTTGAGACCGGACCCCTCATCTCTCAT	120	


```
QY 2281 TCCACACGAGGCTCTCCAGAGCCACCCCATGGAGGACATCGGGCAGGTCTCCCTGGT 2340
Db |||||
QY 2730 TCCACACGAGGCTCTCCAGAGCCACCCCATGGAGGACATCGGGCAGGTCTCCCTGGT 2789
Db |||||
QY 2341 GCAGATCCGGACCTGTCCAGTGATGCGAGGGCGGACACAAATATCCCTGGACATTTTC 2400
Db |||||
QY 2790 GCAGATCCGGACCTGTCCAGTGATGCGAGGGCGGACACAAATATCCCTGGACATTTTC 2849
Db |||||
QY 2401 CGAGGTGGACCCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db |||||
QY 2850 CGAGGTGGACCCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2909
Db |||||
QY 2461 CGAGTTTATGATCTTCAGGAAGTCCCAAGTCCGCTCAGGCAGAGCGCCCTCCCCAT 2520
Db |||||
QY 2910 CGAGTTTATGATCTTCAGGAAGTCCCAAGTCCGCTCAGGCAGAGCGCCCTCCCCAT 2969
Db |||||
QY 2521 GGCTGAGAGAGGTGCGCCAGTTCGCGAGCCACAGTGGCCCTGGCCAGGTGAATGGG 2580
Db |||||
QY 2970 GGCTGAGAGAGGTGCGCCAGTTCGCGAGCCACAGTGGCCCTGGCCAGGTGAATGGG 3029
Db |||||
QY 2581 CCCCCACGAGGCTGGAGATCACAGAGAGTCAAGAGATGTGGAACGCTGTGGCAGA 2640
Db |||||
QY 3030 CCCCCACGAGGCTGGAGATCACAGAGAGTCAAGAGATGTGGAACGCTGTGGCAGA 3089
Db |||||
QY 2641 GGCTGCCGTGGCAGGAAGCGCAAGTGGTCTCGCCCTCAGCAGCTCTTCCACTTCCC 2700
Db |||||
QY 3090 GGCTGCCGTGGCAGGAAGCGCAAGTGGTCTCGCCCTCAGCAGCTCTTCCACTTCCC 3149
Db |||||
QY 2701 TGGAGGACCTGCGCTGGATGAGCTGCAGAGCTGGGGCTGGTGAGAGTGAAGGC 2760
Db |||||
QY 3150 TGGAGGACCTGCGCTGGATGAGCTGCAGAGCTGGGGCTGGTGAGAGTGAAGGC 3209
Db |||||
QY 2761 CTCGCTGGAGCACATCTCCCGGATCTTGAAGGGCAGGCCGGAAGTCTGGAGAGGAGG 2820
Db |||||
QY 3210 CTCGCTGGAGCACATCTCCCGGATCTTGAAGGGCAGGCCGGAAGTCTGGAGAGGAGG 3269
Db |||||
QY 2821 GCGCCCGCAGGAAGCAGGCTTGTCTTCCGCTCTCAGGTCTGAAGAGCTGGGA 2880
Db |||||
QY 3270 GCGCCCGCAGGAAGCAGGCTTGTCTTCCGCTCTCAGGTCTGAAGAGCTGGGA 3329
Db |||||
QY 2881 CCAGAGCCGACATCTCTAAGGAGCTCTCAGATGAGACTGTGTCTCTGGGCCAGTCACT 2940
Db |||||
QY 3330 CCAGAGCCGACATCTCTAAGGAGCTCTCAGATGAGACTGTGTCTCTGGGCCAGTCACT 3389
Db |||||
QY 2941 GACACTGGCTGCGCAGGTGTGAGCCAGCCAGCTGCCAGGCCACCTGGAGCAAGACGG 3000
Db |||||
QY 3390 GACACTGGCTGCGCAGGTGTGAGCCAGCCAGCTGCCAGGCCACCTGGAGCAAGACGG 3449
Db |||||
QY 3001 A 3001
Db 3450 A 3450
```

RESULT 2

```
US-10-077-130-3
; Sequence 3, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7893
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-077-130-3
```

```
Query Match 100.0%; Score 3001; DB 13; Length 7893;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CTTACGAGCAGAACTCGGSCCCAGGCATACAGGAGCGAGACATCTCGCCGCGCTGAG 60
Db |||||
QY 3486 CTTACGAGCAGAACTCGGSCCCAGGCATACAGGAGCGAGACATCTCGCCGCGCTGAG 3545
Db |||||
QY 61 CCACCCGCTGCTCAACCGGGCTGTGGACCACTTTGAGACCGGCAAGACCTCTCATCTCAT 120
Db |||||
QY 3546 CCACCCGCTGCTCAACCGGGCTGTGGACCACTTTGAGACCGGCAAGACCTCTCATCTCAT 3605
Db |||||
QY 121 CTTGAGCTGTGCTCATCTCCAGAGAGCTGTGGAACCGCTCTACAGGAAGCGCGTGTGAC 180
Db |||||
QY 3606 CTTGAGCTGTGCTCATCTCCAGAGAGCTGTGGAACCGCTCTACAGGAAGCGCGTGTGAC 3665
Db |||||
QY 181 GGAGGCGGAGGTCAAGGTCTACATCCAGCAGCTGGTGGAGGGCTGCACCTACCTGCACAG 240
Db |||||
QY 3666 GGAGGCGGAGGTCAAGGTCTACATCCAGCAGCTGGTGGAGGGCTGCACCTACCTGCACAG 3725
Db |||||
QY 241 CCATGGCGTTCTCACCTGGACATAAAGCCCTCTAAACATCTCTGATGGTGCATCTCTGCCG 300
Db |||||
QY 3726 CCATGGCGTTCTCACCTGGACATAAAGCCCTCTAAACATCTCTGATGGTGCATCTCTGCCG 3785
Db |||||
QY 301 GGAAGACATTTAAATCTGCGACTTTTGGCTTTTGGCCAGAACATCAACCCAGCAGAGCTGCA 360
Db |||||
QY 3786 GGAAGACATTTAAATCTGCGACTTTTGGCTTTTGGCCAGAACATCAACCCAGCAGAGCTGCA 3845
Db |||||
QY 361 GTTCAGCAGTACGGCTCCCTCAGTTCTGCTCCCGGAGATCATCCAGCAGAACCTCTGT 420
Db |||||
QY 3846 GTTCAGCAGTACGGCTCCCTCAGTTCTGCTCCCGGAGATCATCCAGCAGAACCTCTGT 3905
Db |||||
QY 421 GAGCAGAGCTCCGACATTTGGGCGCATGGGTGTCTCTCTACCTCAGCTCAGCTGCTC 480
Db |||||
QY 3906 GAGCAGAGCTCCGACATTTGGGCGCATGGGTGTCTCTCTACCTCAGCTCAGCTGCTC 3965
Db |||||
QY 481 ATCCCAATTTGCGGCGAGAGTGAACGCTGCGCACCTCTCTGAAAGCTCTGAGAGGGCGGT 540
Db |||||
QY 3966 ATCCCAATTTGCGGCGAGAGTGAACGCTGCGCACCTCTCTGAAAGCTCTGAGAGGGCGGT 4025
Db |||||
QY 541 GTCATGAGCAGAGCCCATCTGCTGCGCACCTCAGGGAAGAGCGCCAAAGACTTCATCAAGGC 600
Db |||||
QY 4026 GTCATGAGCAGAGCCCATCTGCTGCGCACCTCAGGGAAGAGCGCCAAAGACTTCATCAAGGC 4085
Db |||||
QY 601 TAGCTGSCAGAGAGCCCTCAGAGCCGCGCTAGTGGCGCCAGTGCCTCTCCCAACCCCTG 660
Db |||||
QY 4086 TAGCTGSCAGAGAGCCCTCAGAGCCGCGCTAGTGGCGCCAGTGCCTCTCCCAACCCCTG 4145
Db |||||
QY 661 GTTCTCTGAAATCCATGCTCGGAGGAGGCCACTTTCATCAACCAAGCAGCTCAAGTT 720
Db |||||
QY 4146 GTTCTCTGAAATCCATGCTCGGAGGAGGCCACTTTCATCAACCAAGCAGCTCAAGTT 4205
Db |||||
QY 721 CTTCTGGCCCGAAGTCTGCTGGCAGAGTTCCTCTGATGAGCTCAAGTTCATCTCTGGTGTAT 780
Db |||||
QY 4206 CTTCTGGCCCGAAGTCTGCTGGCAGAGTTCCTCTGATGAGCTCAAGTTCATCTCTGGTGTAT 4265
Db |||||
QY 781 GCGCTCCATCCCTGAGCTGTGCGGGGCCACCCAGCAGCCCTCCCTCGGCGTAGGCCG 840
Db |||||
QY 4266 GCGCTCCATCCCTGAGCTGTGCGGGGCCACCCAGCAGCCCTCCCTCGGCGTAGGCCG 4325
Db |||||
QY 841 GCACCTCTGAGGAGACACTGCTGGTGGCTCTCTCAGTTCTCTCTCTCTCTGACAAAGAGCT 900
Db |||||
QY 4326 GCACCTCTGAGGAGACACTGCTGGTGGCTCTCTCAGTTCTCTCTCTCTCTGACAAAGAGCT 4385
Db |||||
QY 901 CGCCCAATTTGCCGCGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACCACCTGCT 960
Db |||||
QY 4386 CGCCCAATTTGCCGCGGCTAAGTCACTGCCACCTCTCCCGGTGACACACTCACCACCTGCT 4445
Db |||||
QY 961 GCACCCCGGGGCTTCTCTGCGGCCCTCGGCGAGCTGCGCTGAGGAAGCCGAGGCGCAGTGA 1020
Db |||||
```

Db 4446 GCACCCCGGGGCTTCTCGGCGCCCTCGGCGAGCCTGCTGAGGAAGCCGAGGCCAGTGA 4505
Qy 1021 GCGCTCCACGAGGCCCCAGGCTCCGCTCGCATCTCCGAGGCTGCGGGCCACCGGGCGC 1080
Db 4506 GCGCTCCACGAGGCCCCAGGCTCCGCTCGCATCTCCGAGGCTGCGGGCCACCGGGCGC 4565
Qy 1081 CCAGGGCTGCGTGCCTCCGCGCACAGGCTCATCCGAGCCTGTCTACCAACGAGGGGGTGA 1140
Db 4566 CCAGGGCTGCGTGCCTCCGCGCACAGGCTCATCCGAGCCTGTCTACCAACGAGGGGGTGA 4625
Qy 1141 GAGCCCTGAGCACGCGGGCCCTGGCCCCGCGGAGCAGGGCCACCGGGCCCGGGCGGCA 1200
Db 4626 GAGCCCTGAGCACGCGGGCCCTGGCCCCGCGGAGCAGGGCCACCGGGCCCGGGCGGCA 4685
Qy 1201 CTCTGTGAAGGCGGGTACATTTGCGGGGCGCTGCGAGGCTGCGCGAGCCACTGATGGA 1260
Db 4686 CTCTGTGAAGGCGGGTACATTTGCGGGGCGCTGCGAGGCTGCGCGAGCCACTGATGGA 4745
Qy 1261 GCACCGCTGTGAGGAGGAGGCGCGCAGGAGGAGCAGGCCACCTCTCTGGGCCAAGC 1320
Db 4746 GCACCGCTGTGAGGAGGAGGCGCGCAGGAGGAGCAGGCCACCTCTCTGGGCCAAGC 4805
Qy 1321 CCCCTCATTTGAGACTGCGCTCGGCTGCGCTGCGGAGCACCCCGGCCCTCTCGAGGCGCTGCGG 1380
Db 4806 CCCCTCATTTGAGACTGCGCTCGGCTGCGCTGCGGAGCACCCCGGCCCTCTCGAGGCGCTGCGG 4865
Qy 1381 CAGCACTCTCTGAAACATGACTCTCGGAGCACCCCGGCCCTCTCGAGGCGCTGCGG 1440
Db 4866 CAGCACTCTCTGAAACATGACTCTCGGAGCACCCCGGCCCTCTCGAGGCGCTGCGG 4925
Qy 1441 TGAGGCAACGAGCTGCTTTCAGGCGGCTTTCAGGCGGCGGCGGCGGCTTTCAGGCAACATGAGGCA 1500
Db 4926 TGAGGCAACGAGCTGCTTTCAGGCGGCTTTCAGGCGGCGGCGGCGGCTTTCAGGCAACATGAGGCA 4985
Qy 1501 CCCTCAGGGCTCCAGCAGCTTTCATCTGCTGCGGAGCACCCCGGCCCTCTCGAGGCGCTGCGG 1560
Db 4986 CCCTCAGGGCTCCAGCAGCTTTCATCTGCTGCGGAGCACCCCGGCCCTCTCGAGGCGCTGCGG 5045
Qy 1561 GAGGCCATTCCTCGGAGCAGCTTTCAGGCGGAGCACCCCGGCCCTCTCGAGGCGCTGCGG 1620
Db 5046 GAGGCCATTCCTCGGAGCAGCTTTCAGGCGGAGCACCCCGGCCCTCTCGAGGCGCTGCGG 5105
Qy 1621 TTCTGCCCCCAGAGAGGCTGCGAGCCCCCAGCAGGAGTTCGCCCCATGCGCTTCTGCGGCA 1680
Db 5106 TTCTGCCCCCAGAGAGGCTGCGAGCCCCCAGCAGGAGTTCGCCCCATGCGCTTCTGCGGCA 5165
Qy 1681 CTTCCCTCCAGGATCTTTGAAAGAGGCGCCCTTAGTACCCTCAAGCCCCCTTCTTGGGACA 1740
Db 5166 CTTCCCTCCAGGATCTTTGAAAGAGGCGCCCTTAGTACCCTCAAGCCCCCTTCTTGGGACA 5225
Qy 1741 GCCCCAGGACCCCTGCGCCCTGCGCAAGCAAGCCCCCGCATGGAATCTTAAGATGGGGC 1800
Db 5226 GCCCCAGGACCCCTGCGCCCTGCGCAAGCAAGCCCCCGCATGGAATCTTAAGATGGGGC 5285
Qy 1801 TGGAGACATCTCTTCTCGGAGGCAAAACCGGCGCTGAGTTCCCGAGGTCAGC 1860
Db 5286 TGGAGACATCTCTTCTCGGAGGCAAAACCGGCGCTGAGTTCCCGAGGTCAGC 5345
Qy 1861 CTCCAGGCGAGCTCTTCCAGTGAAGTCTCTCAGGCTGGGCTCTCCAGGTCGGGAC 1920
Db 5346 CTCCAGGCGAGCTCTTCCAGTGAAGTCTCTCAGGCTGGGCTCTCCAGGTCGGGAC 5405
Qy 1921 AGAGCTGGCCCTCTCTGAGTGGGAGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 1980
Db 5406 AGAGCTGGCCCTCTCTGAGTGGGAGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 5465
Qy 1981 CTCACACCCACCTTCAGCGGCGCTCAGGAACAGGCGACATGCGCAAGTCTTCCCTGGG 2040
Db 5466 CTCACACCCACCTTCAGCGGCGCTCAGGAACAGGCGACATGCGCAAGTCTTCCCTGGG 5525
Qy 2041 TGTTCGGGGGGTACGCAAGGCTGCTGGCTATGGCACTTTGCTTTGGTGGAGATGC 2100
Db 5526 TGTTCGGGGGGTACGCAAGGCTGCTGGCTATGGCACTTTGCTTTGGTGGAGATGC 5585

Qy 2101 AGGGGGCATCTCTGGGCGAGGGGCCCATGTGGGCCAGGATAGCCTGGGCTGTGCCAGTC 2160
Db 5586 AGGGGGCATCTCTGGGCGAGGGGCCCATGTGGGCCAGGATAGCCTGGGCTGTGCCAGTC 5645
Qy 2161 GAGGAGGAGGAGCAGGAGGAGGCCAGGGCTGAGTCCAGTTCGAGGAGCAGCAGGAGGC 2220
Db 5646 GAGGAGGAGGAGCAGGAGGAGGCCAGGGCTGAGTCCAGTTCGAGGAGCAGCAGGAGGC 5705
Qy 2221 CAGGGCTGAGAGGCCCATCTGCCCCAGGTCAGTGCAGGCGCTGTGCTGAGGTCGGCAGGC 2280
Db 5706 CAGGGCTGAGAGGCCCATCTGCCCCAGGTCAGTGCAGGCGCTGTGCTGAGGTCGGCAGGC 5765
Qy 2281 TCCACACAGGAGCTCTCCAGAGCCACCCATCGGAGGACATCGGCGAGCTCTCCCTGGT 2340
Db 5766 TCCACACAGGAGCTCTCCAGAGCCACCCATCGGAGGACATCGGCGAGCTCTCCCTGGT 5825
Qy 2341 GCAGATCCGGGACCTCTCAGGTCAGGTCGAGGCGCGCGACACAATATCCCTGGACATTC 2400
Db 5826 GCAGATCCGGGACCTCTCAGGTCAGGTCGAGGCGCGCGACACAATATCCCTGGACATTC 5885
Qy 2401 CGAGGTGGACCCCGCCCTACTCAACTCTCAGACTGTAGATATCAAGTACCTCCCAT 2460
Db 5886 CGAGGTGGACCCCGCCCTACTCAACTCTCAGACTGTAGATATCAAGTACCTCCCAT 5945
Qy 2461 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCCGCCCTCCCCCAT 2520
Db 5946 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCCGCCCTCCCCCAT 6005
Qy 2521 GGTGAGGAGGAGCTGGCGGAGTTCCCGAGCCCAAGTGGCCCTGGCCAGGTGAATGGG 2580
Db 6006 GGTGAGGAGGAGCTGGCGGAGTTCCCGAGCCCAAGTGGCCCTGGCCAGGTGAATGGG 6065
Qy 2581 CCCCCAGCAGGCTGGAGATCACAGAGAGTCAGAGATGTGGACGCGCTGCTGGCAGA 2640
Db 6066 CCCCCAGCAGGCTGGAGATCACAGAGAGTCAGAGATGTGGACGCGCTGCTGGCAGA 6125
Qy 2641 GGTGCGCTGGGCGAGGAAGCGCAAGTGTCTCGCGCTCACGCAGCCTCTTCCACTTCCC 2700
Db 6126 GGTGCGCTGGGCGAGGAAGCGCAAGTGTCTCGCGCTCACGCAGCCTCTTCCACTTCCC 6185
Qy 2701 TGGAGGACCTTGGCGCTGGATGAGCTGCAGAGCTGGGGCTGCTGAGAGTGAAGGC 2760
Db 6186 TGGAGGACCTTGGCGCTGGATGAGCTGCAGAGCTGGGGCTGCTGAGAGTGAAGGC 6245
Qy 2761 CTCCGCTGAGCAGCATCTCCCGGATCTGAAGGCGAGGCCGGAAGTCTGGGAAGGAGG 2820
Db 6246 CTCCGCTGAGCAGCATCTCCCGGATCTGAAGGCGAGGCCGGAAGTCTGGGAAGGAGG 6305
Qy 2821 GCCCCCAGGAAGAAAGCAGGCTTGTCTTCCGCTCTCAGGCTCTGAAGAGCTGGGA 2880
Db 6306 GCCCCCAGGAAGAAAGCAGGCTTGTCTTCCGCTCTCAGGCTCTGAAGAGCTGGGA 6365
Qy 2881 CCGAGGCGGACATCTTCAAGGAGCTCAGAGAGCTCAGATGAGACTGTGGCTGGGCGAGT 2940
Db 6366 CCGAGGCGGACATCTTCAAGGAGCTCAGATGAGACTGTGGCTGGGCGAGT 6425
Qy 2941 GACACTGGCCTGCCAGGTGTGAGCCAGCAGCTGCCAGGCGACCTGAGCAAAAGACGG 3000
Db 6426 GACACTGGCCTGCCAGGTGTGAGCCAGCAGCTGCCAGGCGACCTGAGCAAAAGACGG 6485
Qy 3001 A 3001
Db 6486 A 6486

RESULT 3

US-10-077-130-1

; Sequence 1, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.

TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family

TITLE OF INVENTION: Members and Uses Therefor

FILE REFERENCE: MPI2001-047P1RCP1 (M)

CURRENT APPLICATION NUMBER: US/10/07

CURRENT FILING DATE: 2002-02-15
PRIOR APPLICATION NUMBER: 60/2689201

PRIOR APPLICATION NUMBER: 60/
PRIOR FILING DATE: 2001-02-15

NUMBER OF SEQ ID NOS: 9

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 1

LENGTH: 8106

TYPE: DNA
ORGANISM:

ORGANISM: Homo sapiens
FEATURE:

FEATURE:
NAME/KEY:

NAME/KEY: 3 UIR
LOCATION: (1)... (71)

NAME/KEY: CDS

LOCATION: (72) ... (79)

NAME/KEY: 3' UTR

LOCATION: (7965) ... (810

S-10-077-130-1

Query Match

Query Match

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3001; Conservative 0; Mismatches

1	Qy	CCTACGAGCAGAACTCGGGCCCAAGGATACAGGAGGAGGAGACATCTGCGCGCGCTGAG	60
3557	Db		3616
	Db	CCTACGAGCAGAACTCGGGCCCAAGGATACAGGAGGAGGAGACATCTGCGCGCGCTGAG	
61	Qy	CCACCCGCTGCTCAGCGGGCTGCTGGACCACTGTTTGAGACCGCAAGACCCCTCATCCTCAT	120
	Db		
3617	Db	CCACCCGCTGCTCAGCGGGCTGCTGGACCACTGTTTGAGACCGCAAGACCCCTCATCCTCAT	3676
	Qy		
121	Qy	CCTGGAGCTGTGCTCATCCGAGGAGCTGCTGGACCGGCTGTATACAGGAAGGCGCTGGTGAC	180
	Db		
3677	Db	CCTGGAGCTGTGCTCATCCGAGGAGCTGCTGGACCGGCTGTATACAGGAAGGCGCTGGTGAC	3736
	Qy		
181	Qy	GGAGGCCGAGGTCGAAGTCTATATCCAGCAGCTGGTGGAGGGGCTGCACTACCTGCAAG	240
	Db		
3737	Db	GGAGGCCGAGGTCGAAGTCTATATCCAGCAGCTGGTGGAGGGGCTGCACTACCTGCAAG	3796
	Qy		
241	Db	CCATGGGCTTCTCCACCTGGACATAAAGCCCTCTAAATCCTGATGCTGATCTCTGCCG	300
	Db		
3797	Db	CCATGGGCTTCTCCACCTGGACATAAAGCCCTCTAAATCCTGATGCTGATCTCTGCCG	3856
	Qy		
301	Db	GGAAAGACATTAAATCTGCGACTTTGGCTTTGGCCAGAAATACACCCAGCAGAGCTGCA	360
	Db		
3857	Db	GGAAAGACATTAAATCTGCGACTTTGGCTTTGGCCAGAAATACACCCAGCAGAGCTGCA	3916
	Qy		
361	Db	GTTTCAGCCAGTACGGCTCCCTCGATGTTCTGCTCCCGCGAGATCATCCAGCAGAACCCCTGT	420
	Db		
3917	Db	GTTTCAGCCAGTACGGCTCCCTCGATGTTCTGCTCCCGCGAGATCATCCAGCAGAACCCCTGT	3976
	Qy		
421	Db	GAGCGAAGCCTCCGACATTTTGGGCCATGGGTGTCTCTCTACCTCAGCTGACCTGCTC	480
	Db		
3977	Db	GAGCGAAGCCTCCGACATTTTGGGCCATGGGTGTCTCTCTCTACCTCAGCTGACCTGCTC	4036
	Qy		
481	Db	ATCCCCATTTGCGCGCGAGAGTGACCGTGCCACCCTCTTGAACTGCTGGAGGGGCGGT	540
	Db		
4037	Db	ATCCCCATTTGCGCGCGAGAGTGACCGTGCCACCCTCTTGAACTGCTGGAGGGGCGGT	4096
	Qy		
541	Db	GTCATGAGCAGCCCATGGCTGCCACCTCAGCGGAAGAGCCCAAGACTTTCATCAAGGC	600
	Db		
4097	Db	GTCATGAGCAGCCCATGGCTGCCACCTCAGCGGAAGAGCCCAAGACTTTCATCAAGGC	4156
	Qy		
601	Db	TACGCTGCAGAGAGCCCTCATAGGCCCGGCTAGTGGCGGCCAGGTGCCTCTCTCCACCCCTG	660
	Db		
4157	Db	TACGCTGCAGAGAGCCCTCATAGGCCCGGCTAGTGGCGGCCAGGTGCCTCTCTCCACCCCTG	4216
	Qy		
661	Db	GTTCTGAAATCCATGCTCGGAGGAGGGCCATTATCATAAACCAAGAGAGCTCAAGTT	720
	Db		
4217	Db	GTTCTGAAATCCATGCTCGGAGGAGGGCCATTATCAACCAAGAGAGCTCAAGTT	4276
	Qy		

QY 1801 TGGAGACATCTCTCTCTGGGAGGCAAAACCCGGCCCCCTGCGATTTCCCAAGGTCAGC 1860
Db 5357 TGGAGACATCTCTCTCTGGGAGGCAAAACCCGGCCCCCTGCGATTTCCCAAGGTCAGC 5416
QY 1861 CTCCACGAGGAGCTCTTCCCAAGTCAGCTCCCTCAGGCTGGCTCCTCCAGAGTGGGCAC 1920
Db 5417 CTCCACGAGGAGCTCTTCCCAAGTCAGCTCCCTCAGGCTGGCTCCTCCAGAGTGGGCAC 5476
QY 1921 AGAGCTGGCCCCCTCTCTGGATGCGAGGGCTGGAACCCAGGAGGCTGAGGATCTGTCCGA 1980
Db 5477 AGAGCTGGCCCCCTCTCTGGATGCGAGGGCTGGAACCCAGGAGGCTGAGGATCTGTCCGA 5536
QY 1981 CTCACACCCACCTTGACAGGCTCTCAGAACAGGCGAACCATGCGCAAGTTCTCCCTGGG 2040
Db 5537 CTCACACCCACCTTGACAGGCTCTCAGAACAGGCGAACCATGCGCAAGTTCTCCCTGGG 5596
QY 2041 TGGTCGGGGGGCTACGACAGGCGTGGCTATGCGACCTTTGCGTCTTGGTGAGATGC 2100
Db 5597 TGGTCGGGGGGCTACGACAGGCGTGGCTATGCGACCTTTGCGTCTTGGTGAGATGC 5656
QY 2101 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGCCAGTC 2160
Db 5657 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGCCAGTC 5716
QY 2161 GGAGGAGGAGCAGGAGGAGGCGAGGCTGAGTCCAGTCGGAGGAGCAGAGGAGGC 2220
Db 5717 GGAGGAGGAGCAGGAGGAGGCGAGGCTGAGTCCAGTCGGAGGAGCAGAGGAGGC 5776
QY 2221 CAGGCTGAGAGCCACCTGCCCCAGCTCAGTGCAGAGGCTGTGCTCAGAGTCGCGAGGCG 2280
Db 5777 CAGGCTGAGAGCCACCTGCCCCAGCTCAGTGCAGAGGCTGTGCTCAGAGTCGCGAGGCG 5836
QY 2281 TCCACACGAGAGCTCTCAGAGGCCACCCCATGAGGAGACATCGGGCAGAGTCTCCCTGGT 2340
Db 5837 TCCACACGAGAGCTCTCAGAGGCCACCCCATGAGGAGACATCGGGCAGAGTCTCCCTGGT 5896
QY 2341 GCAGATCGGAGCTGTCCAGTGTATGCGAGGCGGCGCACAAATATCCCTGGACATTC 2400
Db 5897 GCAGATCGGAGCTGTCCAGTGTATGCGAGGCGGCGCACAAATATCCCTGGACATTC 5956
QY 2401 CGAGGTGAGACCCCGCTACCTCAACCTCTCAGACCTGTGAGATATCAAGTACCTCCCAT 2460
Db 5957 CGAGGTGAGACCCCGCTACCTCAACCTCTCAGACCTGTGAGATATCAAGTACCTCCCAT 6016
QY 2461 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCGCTCAGCAGAGCGCCCTCCCCAT 2520
Db 6017 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCGCTCAGCAGAGCGCCCTCCCCAT 6076
QY 2521 GGCTGAGGAGAGCTGGCCGAGTTCCCGAGGCCACGTCGCTGCGCAGTGAATCGG 2580
Db 6077 GGCTGAGGAGAGCTGGCCGAGTTCCCGAGGCCACGTCGCTGCGCAGTGAATCGG 6136
QY 2581 CCCCCACGAGGCTGGAGATCAAGAGGAGTCAGAGGATGTGGAACGCTGCTGGCAGA 2640
Db 6137 CCCCCACGAGGCTGGAGATCAAGAGGAGTCAGAGGATGTGGAACGCTGCTGGCAGA 6196
QY 2641 GGCTGCGTGGCAGGAGCGAAGTGGTCTCTGCTCCTCAGCAGGCTCTTCCACTTCCC 2700
Db 6197 GGCTGCGTGGCAGGAGCGAAGTGGTCTCTGCTCCTCAGCAGGCTCTTCCACTTCCC 6256
QY 2701 TGGGAGGACCTGCGCGTGTGATGAGCTGACAGAGCTGGGGCTCGGTGAGAGTGAAGGC 2760
Db 6257 TGGGAGGACCTGCGCGTGTGATGAGCTGACAGAGCTGGGGCTCGGTGAGAGTGAAGGC 6316
QY 2761 CTCGTGGAGCAATCTCCCGATCTCTGAAGGGCAGGCGCGAAGGTCGTGAGAGGAGG 2820
Db 6317 CTCGTGGAGCAATCTCCCGATCTCTGAAGGGCAGGCGCGAAGGTCGTGAGAGGAGG 6376
QY 2821 GCGCCCGAGGAAGAGCGGCTTGTCTTCTCCGCTCTCAGGCTCTGAAGCTGGGA 2880
Db 6377 GCGCCCGAGGAAGAGCGGCTTGTCTTCTCCGCTCTCAGGCTCTGAAGCTGGGA 6436
QY 2881 CCAGGCGCCGACATTCCTAAGGAGGCTCTCAGATGAGAGTGTGGTCTCTGGGCCAGTC 2940

Db 6437 CCGAGCGCCGACATTCCTAAGGAGCTCTCAGATGAGACTGTGTCTCTGGCCAGTCAGT 6496
QY 2941 GACACTGGCTGCCAGTGTCCAGCCAGCAGCTGCCAGCCACCTGGAGCAAGACGG 3000
Db 6497 GACACTGGCTGCCAGTGTCCAGCCAGCAGCTGCCAGCCACCTGGAGCAAGACGG 6556
QY 3001 A 3001
Db 6557 A 6557
RESULT 4
US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1BCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6
Query Match 100.0%; Score 3001; DB 13; Length 23907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 19500 CCTACGGAGCAGAACTCGGGCCCGAGGCATACAGGAGCGAGACATCTCGCGCGCTGAG 60
Db 19500 CCTACGGAGCAGAACTCGGGCCCGAGGCATACAGGAGCGAGACATCTCGCGCGCTGAG 19559
QY 61 CCACCCGCTGCTCAGGGGCTGCTGGACCAAGTTTGGACCCGCAAGACCTCATCTCAT 120
Db 19560 CCACCCGCTGCTCAGGGGCTGCTGGACCAAGTTTGGACCCGCAAGACCTCATCTCAT 19619
QY 121 CTTGGAGCTGTCTCATCCGAGAGCTGCTGACCGCTGTACAGGAAGGCGGTGTGAC 180
Db 19620 CTTGGAGCTGTCTCATCCGAGAGCTGCTGACCGCTGTACAGGAAGGCGGTGTGAC 19679
QY 181 GGAGGCCGAGGTCGAAGGTCTACATCCAGCAGCTGGTGGAGGGGCTGCACTACTGACAG 240
Db 19680 GGAGGCCGAGGTCGAAGGTCTACATCCAGCAGCTGGTGGAGGGGCTGCACTACTGACAG 19739
QY 241 CCATGGGCTTCCACCTGGGACATTAAGCCCTCTTAACATCTGTGATGGTGATCTCTGCCG 300
Db 19740 CCATGGGCTTCCACCTGGGACATTAAGCCCTCTTAACATCTGTGATGGTGATCTCTGCCG 19799
QY 301 GGAAGACATTAATAATCTCGACCTTTGGCTTTGGCCAGAACATCACCCAGCAGAGCTGCA 360
Db 19800 GGAAGACATTAATAATCTCGACCTTTGGCTTTGGCCAGAACATCACCCAGCAGAGCTGCA 19859
QY 361 GTTCAGCCAGTACGCTCCCTGAGTTGCTCTCCCGGAGATCATCCAGCAGAAACCTGT 420
Db 19860 GTTCAGCCAGTACGCTCCCTGAGTTGCTCTCCCGGAGATCATCCAGCAGAAACCTGT 19919
QY 421 GAGCGAAGCTCCGACATTTGGGCCCATGGGTGTATCTTCTTACCTAGCCTGACCTGTCTC 480
Db 19920 GAGCGAAGCTCCGACATTTGGGCCCATGGGTGTATCTTCTTACCTAGCCTGACCTGTCTC 19979
QY 481 ATCCCAATTTGCGCGCAGAGTGAACCGTCCACCTCTCTGAAACGTCCTGGAGGGGCGGT 540
Db 19980 ATCCCAATTTGCGCGCAGAGTGAACCGTCCACCTCTCTGAAACGTCCTGGAGGGGCGGT 20039

QY	541	GTCTATGAGGAGCCCCCATGGCTGCCCACTCAGCGAAGACGGCAAGACTTTCATCAAGGC	600
Db	2040	GTCTATGAGGAGCCCCCATGGCTGCCCACTCAGCGAAGACGGCAAGACTTTCATCAAGGC	20999
QY	601	TACGCTGCAGAGAGCCCCCTCAGGCGCCGCTAGTGCAGCCCAAGTGCCTCTCCACCCCTG	660
Db	20100	TACGCTGCAGAGAGCCCCCTCAGGCGCCGCTAGTGCAGCCCAAGTGCCTCTCCACCCCTG	20159
QY	661	GTCTCTGAAATCCATGCTGCGGAGGAGGCCACCTTCATCAACCAAGCAGCTCAAGTT	720
Db	20160	GTCTCTGAAATCCATGCTGCGGAGGAGGCCACCTTCATCAACCAAGCAGCTCAAGTT	20219
QY	721	CCTCTGCGCCGAGTGCCTGGCAGCGTTCCCTGATGAGCTACAAGTCCATCTCTGGTGAT	780
Db	20220	CCTCTGCGCCGAGTGCCTGGCAGCGTTCCCTGATGAGCTACAAGTCCATCTCTGGTGAT	20279
QY	781	CGCTCTCAATCCCTGAGCTGCTCGGGGGCCACCCGACAGACCCCTCCCTCGGCTAGCCCG	840
Db	20280	CGCTCTCAATCCCTGAGCTGCTCGGGGGCCACCCGACAGACCCCTCCCTCGGCTAGCCCG	20339
QY	841	GCACCTCTGCAAGGACACTGTGTGCTCTCCAGTTCTCTCTCTCTCTGACAAGAGCT	900
Db	20340	GCACCTCTGCAAGGACACTGTGTGCTCTCTCCAGTTCTCTCTCTCTCTGACAAGAGCT	20399
QY	901	CGCCCATTTGCGCGGGTAAGTCACTGCGACCTCCCGGTGACACACTCACCACTGCT	960
Db	20400	CGCCCATTTTCCGCGGGTAAGTCACTGCGACCTCCCGGTGACACACTCACCACTGCT	20459
QY	961	GCACCCCGGGCTTCTCGCGCCCTCGGCAGCGCTGCTGAGGAAGCCGAGGCGCAGTGA	1020
Db	20460	GCACCCCGGGCTTCTCGCGCCCTCGGCAGCGCTGCTGAGGAAGCCGAGGCGCAGTGA	20519
QY	1021	CGCTTCCACGAGGCCACAGCTCCGCTGCTGATCTCCGAGGGTGCCGGGCCACCGGCGC	1080
Db	20520	CGCTTCCACGAGGCCACAGCTCCGCTGCTGATCTCCCGAGGGTGCCGGGCCACCGGCGC	20579
QY	1081	CCAGGGCTGCGTGCCTGCGCGGACAGGGTCACTCGGAGCGCTGTTCTACACAGGCGGGTGA	1140
Db	20580	CCAGGGCTGCGTGCCTGCGCGGACAGGGTCACTCGGAGCGCTGTTCTACACAGGCGGGTGA	20639
QY	1141	GAGCCCTGAGCAGCGGGCCCTGGCCCCGGGAGCAGCGGCAACCGGCCCGCGCGGCA	1200
Db	20640	GAGCCCTGAGCAGCGGGCCCTGGCCCCGGGAGCAGCGGCAACCGGCCCGCGCGGCA	20699
QY	1201	CCTGCTGAAGCGCGGTTACATTTGCGGGGCGCTGCGAGGCTGCGCGAGCCACTGATGGA	1260
Db	20700	CCTGCTGAAGCGCGGTTACATTTGCGGGGCGCTGCGAGGCTGCGCGAGCCACTGATGGA	20759
QY	1261	GCACCGCGTGTGAGGAGGAGGCGCCGAGGGAGCAGGCGCACCTCTCTGGCCCAAGC	1320
Db	20760	GCACCGCGTGTGAGGAGGAGGCGCCGAGGGAGCAGGCGCACCTCTCTGGCCCAAGC	20819
QY	1321	CCCTCTATTGAGACTGCGCTCCGCTGCTGCGACCCCACTTGGGCCCTTGCGCA	1380
Db	20820	CCCTCTATTGAGACTGCGCTCCGCTGCTGCGACCCCACTTGGGCCCTTGCGCA	20879
QY	1381	CAGCCACTCTCTGGAACATGACTCTCCGAGCAGCCCCCGCCCTCTCTCGAGGCGCTCGG	1440
Db	20880	CAGCCACTCTCTGGAACATGACTCTCCGAGCAGCCCCCGCCCTCTCTCGAGGCGCTCGG	20939
QY	1441	TGAGGCA CAGGCACTGCTTACGCCCCCTCGGGGGGGCCCCCTATCAGGGGACATGGGCA	1500
Db	20940	TGAGGCA CAGGCACTGCTTACGCCCCCTCGGGGGGGCCCCCTATCAGGGGACATGGGCA	20999
QY	1501	CCCTCAGGGCTCCAGCAGCTTCCATCCACTGTGGGCCACCCAGGCACTGCTCAGCCAGA	1560
Db	21000	CCCTCAGGGCTCCAGCAGCTTCCATCCACTGTGGGCCACCCAGGCACTGCTCAGCCAGA	21059
QY	1561	GAGGCCATCCCGGACAGCCCTTGGGGGAGCAGCCCTTTCTGCCACCCCAAGCAGGG	1620
Db	21060	GAGGCCATCCCGGACAGCCCTTGGGGGAGCAGCCCTTTCTGCCACCCCAAGCAGGG	21119

QY	1621	TTCTGCCCCCCAGAGGGCTGCAGCCCCCAGCAGAGTTGCCCATGCCCTCTGGCTC	1680
Db	21120	TTCTGCCCCCCAGAGGGCTGCAGCCCCCAGCAGAGTTGCCCATGCCCTCTGGCTC	21179
QY	1681	CTTCCCTCCAGGATCTTTGAAAGAGGGCCCCCTTAGTACCTCTCAAGCCCCCTCTTGGGACA	1740
Db	21180	CTTCCCTCCAGGATCTTTGAAAGAGGGCCCCCTTAGTACCTCTCAAGCCCCCTCTTGGGACA	21239
QY	1741	GCCCCAGGACACCCCTGCTGCTGCAAGAGAGCCCCCATTTGAGCTCTAAGATGGGGCC	1800
Db	21240	GCCCCAGGACACCCCTGCTGCTGCAAGAGAGCCCCCATTTGAGCTCTAAGATGGGGCC	21299
QY	1801	TGGAGACATCTCTCTCTGAGGAGCCAAACCCGCCCCCTGACAGTTCCCGAGGGTCAGC	1860
Db	21300	TGGAGACATCTCTCTCTGAGGAGCCAAACCCGCCCCCTGACAGTTCCCGAGGGTCAGC	21359
QY	1861	CTCCACGCGAGCTCTTCCCAAGTGAGCTCCCTCAGGGTGGGCTCTCCAGGGTGGGAC	1920
Db	21360	CTCCACGCGAGCTCTTCCCAAGTGAGCTCCCTCAGGGTGGGCTCTCCAGGGTGGGAC	21419
QY	1921	AGAGCTTGGCCCCCTCTCTGATGCGAGGGCTGGACCCAGGAGGTGAGGATCTGTCCGA	1980
Db	21420	AGAGCTTGGCCCCCTCTCTGATGCGAGGGCTGGACCCAGGAGGTGAGGATCTGTCCGA	21479
QY	1981	CTCCACACCCACCTTTCAGCGGCTCAGGAAACAGCGGACCATGCGCAAGTTCTCCCTGG	2040
Db	21480	CTCCACACCCACCTTTCAGCGGCTCAGGAAACAGCGGACCATGCGCAAGTTCTCCCTGG	21539
QY	2041	TGGTTCGGGGGGCTTACGAGGCTGGCTGCTATGAGCACTTTGCTTTGGTGGAGATGC	2100
Db	21540	TGGTTCGGGGGGCTTACGAGGCTGGCTGCTATGAGCACTTTGCTTTGGTGGAGATGC	21599
QY	2101	AGGGGGATGCTGGGGCAGGGGCCCATGTGGGCAGAGTACCTGGGCTGTGTCTCCAGTTC	2160
Db	21600	AGGGGGATGCTGGGGCAGGGGCCCATGTGGGCAGAGTACCTGGGCTGTGTCTCCAGTTC	21659
QY	2161	GGAGGAGGAGCAGGAGGAGGAGGCGCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGAGGC	2220
Db	21660	GGAGGAGGAGCAGGAGGAGGAGGCGCAGGGCTGAGTCCAGTCCGAGGAGCAGCAGAGGC	21719
QY	2221	CAGGGCTGAGAGCCACTTGCCTCAGAGCTGAGTGCAGGCTGTGTCTGAGGTCGGCAGGC	2280
Db	21720	CAGGGCTGAGAGCCACTTGCCTCAGAGCTGAGTGCAGGCTGTGTCTGAGGTCGGCAGGC	21779
QY	2281	TCCACACAGGAGCTCTCAGAGCCCCA CCCCATGGGAGGACATCGGGCAGGTCTCCCTGGT	2340
Db	21780	TCCACACAGGAGCTCTCAGAGCCCCA CCCCATGGGAGGACATCGGGCAGGTCTCCCTGGT	21839
QY	2341	GCAGATCCGGACCTGTTCAGGTGATGCGGAGGCGCGACACAAATATCCCTGGACATTTTC	2400
Db	21840	GCAGATCCGGACCTGTTCAGGTGATGCGGAGGCGCGACACAAATATCCCTGGACATTTTC	21899
QY	2401	CGAGGTGAGCCCCCGCTTACCTCAACCTCTCAGAGCTGTACGATATCAAGTACCTCCCAT	2460
Db	21900	CGAGGTGAGCCCCCGCTTACCTCAACCTCTCAGAGCTGTACGATATCAAGTACCTCCCAT	21959
QY	2461	CGAGTTTATGATCTTCAGGAAAGTCCCAAAGTCCGCTCAGCAGAGCGCGCTCCCCCAT	2520
Db	21960	CGAGTTTATGATCTTCAGGAAAGTCCCAAAGTCCGCTCAGCAGAGCGCGCTCCCCCAT	22019
QY	2521	GGCTGAGGAGGAGCTGGCCGAGTTCCCGGAGCCCACTGGGCCCTGGCCAGGTGAACTGG	2580
Db	22020	GGCTGAGGAGGAGCTGGCCGAGTTCCCGGAGCCCACTGGGCCCTGGCCAGGTGAACTGG	22079
QY	2581	CCCCACGCGAGCTGGAGATCACAAGGAGTCAAGAGGATGTGACGCGCTGTCTGGCAGA	2640
Db	22080	CCCCACGCGAGCTGGAGATCACAAGGAGTCAAGAGGATGTGACGCGCTGTCTGGCAGA	22139
QY	2641	GGCTGCCGTGGGAGGAGCGCAAGTGTCTCTCGCGTCAAGCAGCTCTTCCACTTCCC	2700
Db	22140	GGCTGCCGTGGGAGGAGCGCAAGTGTCTCTCGCGTCAAGCAGCTCTTCCACTTCCC	22199
QY	2701	TGGGAGGACACCTCGCGCTGGATGAGCCTGACAGACTGGGGCTGCGTGAGAGGTGAAGGC	2760

Db 22200 TGGAGGACCTGCGCGCTGGATGAGCTTCAGAGCTGGGGCTGCGTGAGAGATGAAGC 22259
Qy 2761 CTCCTGTGAGGACATCTCCCGATCTGAAAGGCGAGCGCGGAAGGTCTGGAGAAGGAGG 2820
Db 22260 CTCCTGTGAGGACATCTCCCGATCTGAAAGGCGAGCGCGGAAGGTCTGGAGAAGGAGG 22319
Qy 2821 GCGCCCGAGGAAGAGCGAGGCTTGCTTCCTTCCGGCTCTCAGGTCTGAAGAGCTGGGA 2880
Db 22320 GCGCCCGAGGAAGAGCGAGGCTTGCTTCCTTCCGGCTCTCAGGTCTGAAGAGCTGGGA 22379
Qy 2881 CCGAGCGCGGAGATTCCTTAAGGAGCTCTCAGATGAGACTGTGCTCTGGGCGAGTCAGT 2940
Db 22380 CCGAGCGCGGAGATTCCTTAAGGAGCTCTCAGATGAGACTGTGCTCTGGGCGAGTCAGT 22439
Qy 2941 GACACTGGCTGCCAGGTGTAGCCCGAGCGCTGCCAGGCGCACTGGAGCAAGAGCGG 3000
Db 22440 GACACTGGCTGCCAGGTGTAGCCCGAGCGCTGCCAGGCGCACTGGAGCAAGAGCGG 22499
Qy 3001 A 3001
Db 22500 A 22500

RESULT 5

US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MP12001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-4

Query Match 100.0%; Score 3001; DB 13; Length 24120;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCTACGAGCAGAACTCGGGCCAGGCATACAGGGAGCGAGACATCTCGGCGCGCTGAG 60
Db 19571 CCTACGAGCAGAACTCGGGCCAGGCATACAGGGAGCGAGACATCTCGGCGCGCTGAG 19630
Qy 61 CCACCCCTGGTACAGGGGCTGCTGGACCAAGTTTGAGACCGGCAAGACCTCATCTCAT 120
Db 19631 CCACCCCTGGTACAGGGGCTGCTGGACCAAGTTTGAGACCGGCAAGACCTCATCTCAT 19690
Qy 121 CCTGGAGCTGTCTATCCAGGAGCTGTGGACCGCTGTACAGGAAGGGCGTGGTGAC 180
Db 19691 CCTGGAGCTGTCTATCCAGGAGCTGTGGACCGCTGTACAGGAAGGGCGTGGTGAC 19750
Qy 181 GGAGCGCAGGTCAAGGTCTACATCCAGAGCTGGTGGAGGGGTGCACTACCTGCACAG 240
Db 19751 GGAGCGCAGGTCAAGGTCTACATCCAGAGCTGGTGGAGGGGTGCACTACCTGCACAG 19810

Qy 241 CCATGGGGTTCTCCACCTGGACATAAAGCCCTTAAACATCTGTGATGTCATCTCTGCCG 300
Db 19811 CCATGGGGTTCTCCACCTGGACATAAAGCCCTTAAACATCTGTGATGTCATCTCTGCCG 19870
Qy 301 GGAAGACATTAAATCTGCGACTTTGGCTTTGCGCCAGAACATCAACCCAGCAGAGCTGCA 360
Db 19871 GGAAGACATTAAATCTGCGACTTTGGCTTTGCGCCAGAACATCAACCCAGCAGAGCTGCA 19930
Qy 361 GTTCAGCAGCTACCGCTCCCTGAGTTTGGCTTCCCGGAGATCATCCAGCAGAACCTGT 420
Db 19931 GTTCAGCAGCTACCGCTCCCTGAGTTTGGCTTCCCGGAGATCATCCAGCAGAACCTGT 19990
Qy 421 GAGCAAGCCTCCGACATTTGGGCAATGGGTGTCTCTTACCTCAGCCTGACTGCTC 480
Db 19991 GAGCAAGCCTCCGACATTTGGGCAATGGGTGTCTCTTACCTCAGCCTGACTGCTC 20050
Qy 481 ATCCCAATTTGCGGCGAGAGTGACCGTGCAACCTCTCTGAAACGTCTCTGAGGGGCGCT 540
Db 20051 ATCCCAATTTGCGGCGAGAGTGACCGTGCAACCTCTCTGAAACGTCTCTGAGGGGCGCT 20110
Qy 541 GTCATGAGCAGCCCATGCTGCCACCTCAGGAAGAGCGCAAGACTTCATCAAGGC 600
Db 20111 GTCATGAGCAGCCCATGCTGCCACCTCAGGAAGAGCGCAAGACTTCATCAAGGC 20170
Qy 601 TACGCTGACAGAGCCCTCAGGCGCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTG 660
Db 20171 TACGCTGACAGAGCCCTCAGGCGCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTG 20230
Qy 661 GTTCTGAAATTCATGCTCGGAGAGGCGCACTTCATCAACCAAGCAGCTCAAGTT 720
Db 20231 GTTCTGAAATTCATGCTCGGAGAGGCGCACTTCATCAACCAAGCAGCTCAAGTT 20290
Qy 721 CTTCTGGGCGGAAAGTGGCTGGCAGGTTCCCTGATGAGCTACAAGTCCATCTCTGGTAT 780
Db 20291 CTTCTGGGCGGAAAGTGGCTGGCAGGTTCCCTGATGAGCTACAAGTCCATCTCTGGTAT 20350
Qy 781 GCGCTCCATCCTGAGCTGTGCGGGGCGCACCCGACAGCCCTCTCTGCGGCTAGCCCG 840
Db 20351 GCGCTCCATCCTGAGCTGTGCGGGGCGCACCCGACAGCCCTCTCTGCGGCTAGCCCG 20410
Qy 841 GCACCTCTGACAGGACACTGGTGGCTCTCTCAGTTCTCTCTCTCTCTGACAAACGAGCT 900
Db 20411 GCACCTCTGACAGGACACTGGTGGCTCTCTCAGTTCTCTCTCTCTCTGACAAACGAGCT 20470
Qy 901 CGCCCAATTTGCGGGGCTAAGTCACTGCACCCCTCCCGGTGACACACTCACACTGCT 960
Db 20471 CGCCCAATTTGCGGGGCTAAGTCACTGCACCCCTCCCGGTGACACACTCACACTGCT 20530
Qy 961 GCAACCCCGGGGCTTCTGCGGCTCTGCGCCAGCTGCTGAGGAAGCCGAGGCGCAGTGA 1020
Db 20531 GCAACCCCGGGGCTTCTGCGGCTCTGCGCCAGCTGCTGAGGAAGCCGAGGCGCAGTGA 20590
Qy 1021 GCGCTCCAGGAGGCGGCTCGGCTGCATCTCCGAGGCTGCGGGGCGCACCGGCGCG 1080
Db 20591 GCGCTCCAGGAGGCGGCTCGGCTGCATCTCCGAGGCTGCGGGGCGCACCGGCGCG 20650
Qy 1081 CCAGGCTCGTGGCGGCGGCTCATCGCAGCTGTTCTACACACAGCGGGTGA 1140
Db 20651 CCAGGCTCGTGGCGGCGGCTCATCGCAGCTGTTCTACACACAGCGGGTGA 20710
Qy 1141 GAGCCTGAGCAGCGGGGCTTCTGCGGCTGCGGAGCAGGCGGCAACCGGCGCGCGCA 1200
Db 20711 GAGCCTGAGCAGCGGGGCTTCTGCGGCTGCGGAGCAGGCGGCAACCGGCGCGCGCA 20770
Qy 1201 CTTGCTGAAGGCGGGTACATTTGCGGGGCGCTGCCAGGCTGCGGAGCCACTGATGA 1260
Db 20771 CTTGCTGAAGGCGGGTACATTTGCGGGGCGCTGCCAGGCTGCGGAGCCACTGATGA 20830
Qy 1261 GCACCGCTGTGAGGAGGAGCGGCCAGGAGGAGGAGGCGCACCTCTGCGCCAAAGC 1320
Db 20831 GCACCGCTGTGAGGAGGAGGCGGCCAGGAGGAGGAGGCGCACCTCTGCGCCAAAGC 20890
Qy 1321 CCCTCATTCGAGACTGCGCTCGGCTGCTGCGACCCACTTGGGCCCTGGGCCA 1380

Db 20891 ||||| CCGCTCATTTGAGACTGCGCTTCCGGCTGCGCTTGGACACCACTTTGGCCCCCTGGCCA 20950
Qy 1381 CAGCACTCCCTGGAACATGACTCTCCGAGACACCCCGCCCTCTCCCTCGAGGGCTTCGG 1440
Db 20951 CAGCACTCCCTGGAACATGACTCTCCGAGACACCCCGCCCTCTCCCTCGAGGGCTTCGG 21010
Qy 1441 TGAGGCACAGGACTGCTTTCAGCCCCCTCCGGGGGGGGCCCCCTATCAGGACATGGGGCA 1500
Db 21011 TGAGGCACAGGACTGCTTTCAGCCCCCTCCGGGGGGGGCCCCCTATCAGGACATGGGGCA 21070
Qy 1501 CCCTCAGGGTCCAAAGCAGCTTCCATCCACTGGTGGCCACCCAGGCACTGCTCAGCCAGA 1560
Db 21071 CCCTCAGGGTCCAAAGCAGCTTCCATCCACTGGTGGCCACCCAGGCACTGCTCAGCCAGA 21130
Qy 1561 GAGGCCATCCCGGACAGCCCTTTGGGGGACCCAGCCCTTTTTCGCCACCCCAAGCAGGG 1620
Db 21131 GAGGCCATCCCGGACAGCCCTTTGGGGGACCCAGCCCTTTTTCGCCACCCCAAGCAGGG 21190
Qy 1621 TTCTGCCCCCNGAGGGCTGCAGCCCCCACCAGCAGTTGGCCCCATGCCCTCTGGCTC 1680
Db 21191 TTCTGCCCCCNGAGGGCTGCAGCCCCCACCAGCAGTTGGCCCCATGCCCTCTGGCTC 21250
Qy 1681 CTTCCCTCCAGGATCTTTCGAAAGAGGCCCTTAGTACCCTCAAGCCCTTCTTTGGGACA 1740
Db 21251 CTTCCCTCCAGGATCTTTCGAAAGAGGCCCTTAGTACCCTCAAGCCCTTCTTTGGGACA 21310
Qy 1741 GCCCCAGGCAACCCCTCCCTGCGCAAGAGCCCGCCCAATGGACTCTAAGATGGGGCC 1800
Db 21311 GCCCCAGGCAACCCCTCCCTGCGCAAGAGCCCGCCCAATGGACTCTAAGATGGGGCC 21370
Qy 1801 TGGAGACATCTCTTCTGGGAGGCCAAACCCGGGCCCTGCAGTTCCCGAGGTCAGC 1860
Db 21371 TGGAGACATCTCTTCTGGGAGGCCAAACCCGGGCCCTGCAGTTCCCGAGGTCAGC 21430
Qy 1861 CTCCAGCGGAGCTCTTCCCAAGTGAAGCTCCCTCAGGGTGGGCTCTCCAGGTGGGCAC 1920
Db 21431 CTCCAGCGGAGCTCTTCCCAAGTGAAGCTCCCTCAGGGTGGGCTCTCCAGGTGGGCAC 21490
Qy 1921 AGAGCTTGGCCCCCTCCCTGGATCGGAGGGCTGGACCCAGAGGCTCAGGATCTGTCCGA 1980
Db 21491 AGAGCTTGGCCCCCTCCCTGGATCGGAGGGCTGGACCCAGAGGCTCAGGATCTGTCCGA 21550
Qy 1981 CTCACACCCACCTTGACGCGGCTCAGGAAACAGGAGACCAATGCGAAGTTCTCCCTGGG 2040
Db 21551 CTCACACCCACCTTGACGCGGCTCAGGAAACAGGAGACCAATGCGAAGTTCTCCCTGGG 21610
Qy 2041 TGGTCGGGGGGCTACGACGCGTGGCTGGCTATGGCACCTTTGCTTGGTGAGATGC 2100
Db 21611 TGGTCGGGGGGCTACGACGCGTGGCTGGCTATGGCACCTTTGCTTGGTGAGATGC 21670
Qy 2101 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCCTGGGCTGTGTCCAGTC 2160
Db 21671 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCCTGGGCTGTGTCCAGTC 21730
Qy 2161 GGAGGAGGAGGAGGAGGAGGAGGAGGCTGAGTCCAGTTCGAGGAGGAGGAGGAGG 2220
Db 21731 GGAGGAGGAGGAGGAGGAGGAGGAGGCTGAGTCCAGTTCGAGGAGGAGGAGGAGG 21790
Qy 2221 CAGGGCTGAGAGGCCACTTGCCCCAGGTCAGTGCAGAGCCCTGTGCTCAGGTCGCGAGGGC 2280
Db 21791 CAGGGCTGAGAGGCCACTTGCCCCAGGTCAGTGCAGAGCCCTGTGCTCAGGTCGCGAGGGC 21850
Qy 2281 TCCACACGAGGAGCTCTCCAGAGCCCAACCCCATGGGAGGACATCGGGCAGGTCCTCCCTGGT 2340
Db 21851 TCCACACGAGGAGCTCTCCAGAGCCCAACCCCATGGGAGGACATCGGGCAGGTCCTCCCTGGT 21910
Qy 2341 GCAGATCCGGAGCTGTGCAGGTGATGCGGAGGCGCGGACACAATATCCCTGGACATTTTC 2400
Db 21911 GCAGATCCGGAGCTGTGCAGGTGATGCGGAGGCGCGGACACAATATCCCTGGACATTTTC 21970
Qy 2401 CGAGGTGAGCCCCGCTTACCTCAGCTCAGACTGTACGATATCAAGTACCTCCCATTT 2460
|||||

Db 21971 CGAGGTGGACCCCGCTTACCTCAACCTCTCAGACCTCTCAGATATCAAGTACCTCCCAT 22030
Qy 2461 CGAGTTTATGATCTTTCAGGAAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 2520
Db 22031 CGAGTTTATGATCTTTCAGGAAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 22090
Qy 2521 GGCTGAGGAGGAGTGGCCGAGTTCCCGGAGCCACAGTGGCCCTGGCCAGGTGAATGGG 2580
Db 22091 GGCTGAGGAGGAGTGGCCGAGTTCCCGGAGCCACAGTGGCCCTGGCCAGGTGAATGGG 22150
Qy 2581 CCCCCAGCAGGCGCTGGAGATCAGAGGAGTTCAGAGGATGTGGACGCGCTCTTCCACTTCCC 2640
Db 22151 CCCCCAGCAGGCGCTGGAGATCAGAGGAGTTCAGAGGATGTGGACGCGCTCTTCCACTTCCC 22210
Qy 2641 GGCTGCCGTGGGAGGAGGAGGAGGAGTGGTCTTCCGCTCAGCAGGCTCTTCCACTTCCC 2700
Db 22211 GGCTGCCGTGGGAGGAGGAGGAGTGGTCTTCCGCTCAGCAGGCTCTTCCACTTCCC 22270
Qy 2701 TGGGAGGACACTCGCCGCTGATGAGCCTCAGAGCTGGGGCTGCTGAGAGGTGAAGGC 2760
Db 22271 TGGGAGGACACTCGCCGCTGATGAGCCTCAGAGCTGGGGCTGCTGAGAGGTGAAGGC 22330
Qy 2761 CTCGCTGGAGCACATCTCCCGGATCTTAAAGGGCAGGCCCGGAAGGTCTGGAGAAAGGAGG 2820
Db 22331 CTCGCTGGAGCACATCTCCCGGATCTTAAAGGGCAGGCCCGGAAGGTCTGGAGAAAGGAGG 22390
Qy 2821 GCCCCCAGAGAAAGCAGGCGCTTGGCTTCTTCCGCTCTCAGGCTCTGAAGAGCTGGGA 2880
Db 22391 GCCCCCAGAGAAAGCAGGCGCTTGGCTTCTTCCGCTCTCAGGCTCTGAAGAGCTGGGA 22450
Qy 2881 CCAGCGCCGACATCTCTAAGGGAGCTCTCAGATGAGACTGTGGTCTTGGGCCAGTCAGT 2940
Db 22451 CCAGCGCCGACATCTCTAAGGGAGCTCTCAGATGAGACTGTGGTCTTGGGCCAGTCAGT 22510
Qy 2941 GACACTGGCTTGCAGGCTGTCCAGCCAGCCAGCTGCCAGGCGCACCTGGAGCAAGAGCG 3000
Db 22511 GACACTGGCTTGCAGGCTGTCCAGCCAGCCAGCTGCCAGGCGCACCTGGAGCAAGAGCG 22570
Qy 3001 A 3001
Db 22571 A 22571

RESULT 6

US-10-307-019-3
; Sequence 3, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DV1
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5007
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (97)...(4926)
US-10-307-019-3

Query Match 99.9%; Score 2999.4; DB 15; Length 5007;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTACGAGCAGAACTCGGGGCCAGGCATACAGGAGCGAGACATCTCTGGCGCGGCTGAG 60
DB CCTACGAGCAGAACTCGGGGCCAGGCATACAGGAGCGAGACATCTCTGGCGCGGCTGAG 581
QY 61 CCACCCGCTGCTACGGGGCTGCTGGACCAAGTTTGAGACCCGCAAGACCTCATCTCTCAT 120
DB CCACCCGCTGCTACGGGGCTGCTGGACCAAGTTTGAGACCCGCAAGACCTCATCTCTCAT 641
QY 121 CCTGAGCTGTGCTCATCCGAGGAGCTCTGGACCGCTGTACAGGAAGGSCGTGTGAC 180
DB CCTGAGCTGTGCTCATCCGAGGAGCTCTGGACCGCTGTACAGGAAGGSCGTGTGAC 701
QY 181 GGAGCCGAGGTCGAAGGTCATACACAGCTGTGTGAGGGGCTGCACTACCTGCAAG 240
DB GGAGCCGAGGTCGAAGGTCATACACAGCTGTGTGAGGGGCTGCACTACCTGCAAG 761
QY 241 CCATGGGCTCTCCACCTGGACATTAAGCCCTCTAAACATCTGTATGTGATCTGTGCCG 300
DB CCATGGGCTCTCCACCTGGACATTAAGCCCTCTAAACATCTGTATGTGATCTGTGCCG 821
QY 301 GGAAGACATTAATAATCTCGCACTTTGGCTTTGGCCAGAAATCAACCCAGCAGAGCTGCA 360
DB GGAAGACATTAATAATCTCGCACTTTGGCTTTGGCCAGAAATCAACCCAGCAGAGCTGCA 881
QY 361 GTTCAGCAGTACGGCTCCCTGAGTTGCTCTCCCGAGATCATCCAGCAGAAACCTGT 420
DB GTTCAGCAGTACGGCTCCCTGAGTTGCTCTCCCGAGATCATCCAGCAGAAACCTGT 941
QY 421 GAGCGAGCTCCGACATTTGGCCATGGGTGTCATCTCTACCTGAGCTGACCTGCTC 480
DB GAGCGAGCTCCGACATTTGGCCATGGGTGTCATCTCTACCTGAGCTGACCTGCTC 1001
QY 481 ATCCCCATTTGCCCGGAGAGTACCGCTGCGACCCCTCTGAACGCTCTGGAGGGGCGCT 540
DB ATCCCCATTTGCCCGGAGAGTACCGCTGCGACCCCTCTGAACGCTCTGGAGGGGCGCT 1061
QY 541 GTCATGAGCAGCCCATGAGTGTGCTGCCACCTGAGGAAAGCGCCAAAGACTTCATCAAGG 600
DB GTCATGAGCAGCCCATGAGTGTGCTGCCACCTGAGGAAAGCGCCAAAGACTTCATCAAGG 1121
QY 601 TACGCTCAGAGAGCCCTCAGGGCCGCGCTAGTGGCGGCCAGTGCCTCTCCACCCCTG 660
DB TACGCTCAGAGAGCCCTCAGGGCCGCGCTAGTGGCGGCCAGTGCCTCTCCACCCCTG 1181
QY 661 GTTCCTGAAATCCATGCTCGGAGGAGGCCATTCATCAACACCAAGCAGCTCAAGTT 720
DB GTTCCTGAAATCCATGCTCGGAGGAGGCCATTCATCAACACCAAGCAGCTCAAGTT 1241
QY 721 CCTCTGGCCGAGTCTGCGCAGCGTTCCGTATGAGCTACAAGTTCATCTCTGTGAT 780
DB CCTCTGGCCGAGTCTGCGCAGCGTTCCGTATGAGCTACAAGTTCATCTCTGTGAT 1241
QY 1242 CCTCTGGCCGAGTCTGCGCAGCGTTCCGTATGAGCTACAAGTTCATCTCTGTGAT 1301
DB CGCTCTCATCCCTGAGCTGTGCGGGGCCACCCGACAGCCCTCTCTGGCGGTAGCCCG 840
DB CGCTCTCATCCCTGAGCTGTGCGGGGCCACCCGACAGCCCTCTCTGGCGGTAGCCCG 1361
QY 841 GCACCTCTGAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTGACAAACAGCT 900
DB GCACCTCTGAGGACACTGGTGGCTCTCCAGTTCTCTCTCTCTCTGACAAACAGCT 1421
QY 901 CGCCCATTTGCGGGGCTTAAGTCACTGCGACCCCTCCCGGTGACACACTCAGCTGT 960
DB CGCCCATTTGCGGGGCTTAAGTCACTGCGACCCCTCCCGGTGACACACTCAGCTGT 1481
QY 961 GCACCCCGGGCTTCTGCGGGCCCTCGGACGCTGCTGAGGAAGCGAGGCGAGTGA 1020
DB GCACCCCGGGCTTCTGCGGGCCCTCGGACGCTGCTGAGGAAGCGAGGCGAGTGA 1541
QY 1021 GCGCTCCACGAGGCCCTCAGCTCCGCTGATCTCCGAGGTTGCGGGCCACCGGCCG 1080
DB GCGCTCCACGAGGCCCTCAGCTCCGCTGATCTCCGAGGTTGCGGGCCACCGGCCG 1601

QY 1081 CCAGGGCTGGTGGTCCCGGACAGCGTCATCCGAGCGCTGTCTTACCAACAGGCGGCTGA 1140
DB CCAGGGCTGGTGGTGGTCCCGGACAGCGTCATCCGAGCGCTGTCTTACCAACAGGCGGCTGA 1661
QY 1141 GAGCCCTGAGCAGCGGGCCCTTGCGCCCGGGGAGCAGCGCGACCCCGCGCGGCGGCA 1200
DB GAGCCCTGAGCAGCGGGCCCTTGCGCCCGGGGAGCAGCGCGACCCCGCGCGGCGGCA 1721
QY 1201 CTTGCTGAAGGGGGGTACATTTGCGGGGGCGCTGCCAGGCGCTGGCGAGCACTGATGA 1260
DB CTTGCTGAAGGGGGGTACATTTGCGGGGGCGCTGCCAGGCGCTGGCGAGCACTGATGA 1781
QY 1261 GCACCGCGTCTCTGAGGAGGAGCGCCAGGAGGAGCAGGCGCAACCTCTCTGCGCAAGC 1320
DB GCACCGCGTCTCTGAGGAGGAGCGCCAGGAGGAGCAGGCGCAACCTCTCTGCGCAAGC 1841
QY 1321 CCCTCATTTGAGACTGCTCCCTCGGCTGCTCTCTGCGACCCCACTTTGGCCCTTGCGCA 1380
DB CCCTCATTTGAGACTGCTCCCTCGGCTGCTCTCTGCGACCCCACTTTGGCCCTTGCGCA 1901
QY 1381 CAGCCACTCCCTGAAACATGACTCTCCGAGCACCCCCCGCCCTCTCTGCGAGGCGCTGCGG 1440
DB CAGCCACTCCCTGAAACATGACTCTCCGAGCACCCCCCGCCCTCTCTGCGAGGCGCTGCGG 1961
QY 1441 TGAGGCACAGCACTGCTCTTCAAGCCCTCTCGGGGGGGCCCTTATCAGGAGCAATGGGGCA 1500
DB TGAGGCACAGCACTGCTCTTCAAGCCCTCTCGGGGGGGCCCTTATCAGGAGCAATGGGGCA 2021
QY 1501 CCTCAGGGCTCCAGCAGCTTCATCTCACTGTGTGGCGACCCAGGCACTGTCTCAGCGAGA 1560
DB CCTCAGGGCTCCAGCAGCTTCATCTCACTGTGTGGCGACCCAGGCACTGTCTCAGCGAGA 2081
QY 1561 GAGGCCATCCCGGACAGCCCTTGCGGGGAGCAGCCCTTTCTGCGCACCCCAAGCAGGG 1620
DB GAGGCCATCCCGGACAGCCCTTGCGGGGAGCAGCCCTTTCTGCGCACCCCAAGCAGGG 2141
QY 1621 TTCTGCCCCCAGGAGGGCTGACAGCCCCCAGCAGTTGCCCCCATGCCCCCTCTGCTGCTC 1680
DB TTCTGCCCCCAGGAGGGCTGACAGCCCCCAGCAGTTGCCCCCATGCCCCCTCTGCTGCTC 2201
QY 1681 CTTCCCTCCAGGATCTTGCAAGAGGCGCCCTTAGTACCTCAAGCCCTTCTTGGGACA 1740
DB CTTCCCTCCAGGATCTTGCAAGAGGCGCCCTTAGTACCTCAAGCCCTTCTTGGGACA 2261
QY 1741 GCCCCAGGCACCCCTCTGCCCCCTGCCAAAGCAAGCCCCCATTTGGACTCTAAAGTTGGGGC 1800
DB GCCCCAGGCACCCCTCTGCCCCCTGCCAAAGCAAGCCCCCATTTGGACTCTAAAGTTGGGGC 2321
QY 1801 TGGAGACATCTCTCTTCTGAGGAGGCCAAAACCCGGCCCCCTGAGTTCCCCAGGCTCAGC 1860
DB TGGAGACATCTCTCTTCTGAGGAGGCCAAAACCCGGCCCCCTGAGTTCCCCAGGCTCAGC 2381
QY 1861 CTTCCAGGCGAGCTCTTCCCAAGTACCTCTCAGGGTGGGCTCTCTCCAGGTGGGCA 1920
DB CTTCCAGGCGAGCTCTTCCCAAGTACCTCTCAGGGTGGGCTCTCTCCAGGTGGGCA 2441
QY 1921 AGAGCTGTGCCCCCTCTCTGATGCGAGGCGCTGAGACCCAGAGGCTGAGGATCTGTCCGA 1980
DB AGAGCTGTGCCCCCTCTCTGATGCGAGGCGCTGAGACCCAGAGGCTGAGGATCTGTCCGA 2501
QY 1981 CTCACACCACTTGTGAGCGGCTCAGAAACAGGCGACCATGCGGAAGTTCTCCCTGGG 2040
DB CTCACACCACTTGTGAGCGGCTCAGAAACAGGCGACCATGCGGAAGTTCTCCCTGGG 2561
QY 2041 TGGTCCGCGGGGCTACGAGGCGTGGCTGATGCGACCTTTGCTTTGGTGGAGATGC 2100
DB TGGTCCGCGGGGCTACGAGGCGTGGCTGATGCGACCTTTGCTTTGGTGGAGATGC 2621
QY 2101 AGGGGCGATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTCTCCAGTTC 2160
DB AGGGGCGATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTCTCCAGTTC 2681
QY 2161 GGAGGAGGAGGAGGAGGAGGCGCCAGGCTGAGTCCAGTCGAGGAGCAGCAGAGGC 2220

QY	841	GCACTCTGACGGGACACTGGTGGCTCTCTCAGTTCTCTCTCTCTGACAAAGAGCT	900
Db	4283	GCACTCTGACGGGACACTGGTGGCTCTCTCAGTTCTCTCTCTCTCTGACAAAGAGCT	4342
QY	901	CGCCCCATTGCGGGCTAAAGTCACTGCGCAACCTCTCCCGGTGACACACTCAACACTGCT	960
Db	4343	CGCCCCATTGCGGGCTAAAGTCACTGCGCAACCTCTCCCGGTGACACACTCAACACTGCT	4402
QY	961	GCAACCCCGGGCTTCTGCGGCCCTCGGCCAGAGCTGCTGAGGAAGCGGAGGCCAGTGA	1020
Db	4403	GCAACCCCGGGCTTCTGCGGCCCTCGGCCAGAGCTGCTGAGGAAGCGGAGGCCAGTGA	4462
QY	1021	GGCTTCCACCGAGGCCCAAGCTTCGCTCGCATCTCCGAGAGGTGCGGGGCCACCGGGCGC	1080
Db	4463	GGCTTCCACCGAGGCCCAAGCTTCGCTCGCATCTCCGAGAGGTGCGGGGCCACCGGGCGC	4522
QY	1081	CCAGGGCTGGCTGCGCGCGGACAGCGTCACTCGGCAAGCTGTTTACACAGAGGGGTGA	1140
Db	4523	CCAGGGCTGGCTGCGCGCGGACAGCGTCACTCGGCAAGCTGTTTACACAGAGGGGTGA	4582
QY	1141	GAGCCCTGAGCAGCGGGGCTTGGCCCCCGGGAGCAGGGCGCACCGGGCCCGGGCGGCA	1200
Db	4583	GAGCCCTGAGCAGCGGGGCTTGGCCCCCGGGAGCAGGGCGCACCGGGCCCGGGCGGCA	4642
QY	1201	CCTGCTGAAGGGCGGGTACATTGCGGGGGCGCTGCGAGGCTGCGCGAGCOACTGATGGA	1260
Db	4643	CCTGCTGAAGGGCGGGTACATTGCGGGGGCGCTGCGAGGCTGCGCGAGCOACTGATGGA	4702
QY	1261	GCACCGCTGCTGAGGAGGAGCGCGCAGGGAGGAGCGGCGCACCGGGCCCGGGCGGCA	1320
Db	4703	GCACCGCTGCTGAGGAGGAGCGCGCAGGGAGGAGCGGCGCACCGGGCCCGGGCGGCA	4762
QY	1321	CCCCTCATTCGAGACTGCCCCCTCGGCTGCTGCTGCGGACCCACCTATGGGCCCTGGCCA	1380
Db	4763	CCCCTCATTCGAGACTGCCCCCTCGGCTGCTGCTGCGGACCCACCTATGGGCCCTGGCCA	4822
QY	1381	CAGCCACTCCCTGGAACATGACTCTCGAGACAACCCCGGCCCTCTCGAGGCGCTCGG	1440
Db	4823	CAGCCACTCCCTGGAACATGACTCTCGAGACAACCCCGGCCCTCTCGAGGCGCTCGG	4882
QY	1441	TGAGGCACAGGCACTGCTTCAGCCCCCTCGGGGGGGCCCCCTATCAGGACACTGGGGCA	1500
Db	4883	TGAGGCACAGGCACTGCTTCAGCCCCCTCGGGGGGGCCCCCTATCAGGACACTGGGGCA	4942
QY	1501	CCCTCAGGGCTCCAAAGCAGCTTCATCCACTGGTGGCCACCCAGGCACTGCTCAGCCAGA	1560
Db	4943	CCCTCAGGGCTCCAAAGCAGCTTCATCCACTGGTGGCCACCCAGGCACTGCTCAGCCAGA	5002
QY	1561	GAGGCCATCCCGGACAGCCCTTGGGGGAGCCAGCCCTTTCTGCCACCCCAAGCAGGG	1620
Db	5003	GAGGCCATCCCGGACAGCCCTTGGGGGAGCCAGCCCTTTCTGCCACCCCAAGCAGGG	5062
QY	1621	TTCTGCCCCCAGAGGGGCTGACGCCCCACCCAGCAGTTTGGCCCCCATGCCCTCTGGCTC	1680
Db	5063	TTCTGCCCCCAGAGGGGCTGACGCCCCACCCAGCAGTTTGGCCCCCATGCCCTCTGGCTC	5122
QY	1681	CTTCCCTCCAGGATCTTTGAAAAGAGGCCCTTTAGTACCTCAAGCCCCCTTTCTGGGACA	1740
Db	5123	CTTCCCTCCAGGATCTTTGAAAAGAGGCCCTTTAGTACCTCAAGCCCCCTTTCTGGGACA	5182
QY	1741	GCCCCAGGCACCCCTGCCCCCTGCCAAAGCAGAGCCCCCATTTGGACTCTAAGATGGGGC	1800
Db	5183	GCCCCAGGCACCCCTGCCCCCTGCCAAAGCAGAGCCCCCATTTGGACTCTAAGATGGGGC	5242
QY	1801	TGAGACATCTCTCTTCTGGGAGGCCAAACCCGGGCCCTCGAGTTCCCAAGGGTCAGC	1860
Db	5243	TGAGACATCTCTCTTCTGGGAGGCCAAACCCGGGCCCTCGAGTTCCCAAGGGTCAGC	5302
QY	1861	CTCCAGGCGAGCTCTTCCCAAGTGAAGTCTCTCAGGGTGGGCTCTCTCCAGAGTGGGCAC	1920
Db	5303	CTCCAGGCGAGCTCTTCCCAAGTGAAGTCTCTCAGGGTGGGCTCTCTCCAGAGTGGGCAC	5362

QY	1921	AGAGCTGGGCCCTCCCTGGATGCGGAGGGCTGACCCAGGAGGCTGAGGATCTGTCCGA	1980
Db	5363	AGAGCTGGGCCCTCCCTGGATGCGGAGGGCTGACCCAGGAGGCTGAGGATCTGTCCGA	5422
QY	1981	CTCCACACCCACCTTGCAGGGGCTCAGGAAACAGGCGACCATGCGCAAGTTCTCCCTGGG	2040
Db	5423	CTCCACACCCACCTTGCAGGGGCTCAGGAAACAGGCGACCATGCGCAAGTTCTCCCTGGG	5482
QY	2041	TGGTCGGGGGGCTACGACAGCGTGGCTGCTATGGCACCTTTGGCTTTGGTGAGATGC	2100
Db	5483	TGGTCGGGGGGCTACGACAGCGTGGCTGCTATGGCACCTTTGGCTTTGGTGAGATGC	5542
QY	2101	AGGGGGCATGCTGGGGCAGGGGCCATGTCGGCCAGGATAGCTGGGCTGTGCCAGTC	2160
Db	5543	AGGGGGCATGCTGGGGCAGGGGCCATGTCGGCCAGGATAGCTGGGCTGTGCCAGTC	5602
QY	2161	GGAGGAGGAGGACGAGGAGGAGGCCAGGGCTGAGTCCCAGTCGAGGAGGAGCAGCAGGAGGC	2220
Db	5603	GGAGGAGGAGGACGAGGAGGAGGCCAGGGCTGAGTCCCAGTCGAGGAGGAGCAGCAGGAGGC	5662
QY	2221	CAGGGCTGAGAGCCCACTGCCCCCAGGTCAAGTGCAGGGCTGTGCTGAGGTTCGGCAGGGC	2280
Db	5663	CAGGGCTGAGAGCCCACTGCCCCCAGGTCAAGTGCAGGGCTGTGCTGAGGTTCGGCAGGGC	5722
QY	2281	TCCACACAGGAGCTCTCCAGAGCCCACTGCGGAGGACATCGGCGAGGTCTCCCTGGT	2340
Db	5723	TCCACACAGGAGCTCTCCAGAGCCCACTGCGGAGGACATCGGCGAGGTCTCCCTGGT	5782
QY	2341	GCAGATCCGGGACCTGTCAAGTGAATCGGAGGGCGGCGCACAAATATCCCTGGACATTC	2400
Db	5783	GCAGATCCGGGACCTGTCAAGTGAATCGGAGGGCGGCGCACAAATATCCCTGGACATTC	5842
QY	2401	CGAGGTGGACCCCGCTACCTCAACCTCTCAGACTGATAGGATCAAGTCACTCCCAT	2460
Db	5843	CGAGGTGGACCCCGCTACCTCAACCTCTCAGACTGATAGGATCAAGTCACTCCCAT	5902
QY	2461	CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCCGCTCCCCCAT	2520
Db	5903	CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCCAGAGCCGCTCCCCCAT	5962
QY	2521	GGCTGAGGAGGAGCTGGCCGAGTTCCGAGGCCACAGTGGCCCTGGCCAGGTGAATCGGG	2580
Db	5963	GGCTGAGGAGGAGCTGGCCGAGTTCCGAGGCCACAGTGGCCCTGGCCAGGTGAATCGGG	6022
QY	2581	CCCCACGACGAGCCTGGAGATCACAGAGAGTCAAGAGGATGAGCGCTGCTGSCAGA	2640
Db	6023	CCCCACGACGAGCCTGGAGATCACAGAGAGTCAAGAGGATGAGCGCTGCTGSCAGA	6082
QY	2641	GGCTGCGTGGGCGAGGAGCGCAAGTGGTCTCGCCGTCACGACGCTTCCACTTCCC	2700
Db	6083	GGCTGCGTGGGCGAGGAGCGCAAGTGGTCTCGCCGTCACGACGCTTCCACTTCCC	6142
QY	2701	TGGGAGGACCTGCGCTGGATGAGCTGAGAGCTGGGGCTGGCTGAGAGAGTGAAGGC	2760
Db	6143	TGGGAGGACCTGCGCTGGATGAGCTGAGAGCTGGGGCTGGCTGAGAGAGTGAAGGC	6202
QY	2761	CTCCCTGGAGCACATCTCCCGGATCCTGAAGGGCAGGCGGAAGTCTTGAAGAGAGG	2820
Db	6203	CTCCCTGGAGCACATCTCCCGGATCCTGAAGGGCAGGCGGAAGTCTTGAAGAGAGG	6262
QY	2821	GGCCCCCAGGAAGAACGACGCGCTTGTCTTCGCTCTCAGGTCTGAAGAGCTGGGA	2880
Db	6263	GGCCCCCAGGAAGAACGACGCGCTTGTCTTCGCTCTCAGGTCTGAAGAGCTGGGA	6322
QY	2881	CGAGGCGCGCATCTTAAGGAGCTCTCAGATGAGACTGTGGTCTGGGGCAGTCAAT	2940
Db	6323	CGAGGCGCGCATCTTAAGGAGCTCTCAGATGAGACTGTGGTCTGGGGCAGTCAAT	6382
QY	2941	GACACTGGCCTGCCAGGTGTGAGCCAGCTGCCCCAGGCGCCACTGAGGCAAGAGCGG	3000
Db	6383	GACACTGGCCTGCCAGGTGTGAGCCAGCTGCCCCAGGCGCCACTGAGGCAAGAGCGG	6442
QY	3001	A 3001	

Db 6443 A 6443

RESULT 8

US-09-858-664A-1

; Sequence 1, Application US/09858664A

; Patent No. US20020072491A1

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui, et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL000927-CIP

; CURRENT APPLICATION NUMBER: US/09/858,664A

; CURRENT FILING DATE: 2001-05-17

; PRIOR APPLICATION NUMBER: 09/711,134

; PRIOR FILING DATE: 2000-11-11

; NUMBER OF SEQ. ID NOS: 33

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 5207

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-858-664A-1

Query Match 99.9%; Score 2997.8; DB 9; Length 5207;

Best Local Similarity 99.9%; Pred. No. 0;

Mismatches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCTACGAGCAGAACTCGGGCCAGGATACAGGAGGAGGAGACATCTTGGCGCGCTGAG 60

Db 699 CCTACGAGCAGAACTCGGGCCAGGATACAGGAGGAGGAGACATCTTGGCGCGCTGAG 758

Qy 61 CCACCCGCTGTCACGGGGCTGCTGGACCACTTTGAGACCGCAAGACCTCATCTCCAT 120

Db 759 CCACCCGCTGTCACGGGGCTGCTGGACCACTTTGAGACCGCAAGACCTCATCTCCAT 818

Qy 121 CCTGAGCTGTCTATCCGAGAGCTGCTGGACCGCTGTACAGGAAGGCGTGGTGAC 180

Db 819 CCTGAGCTGTCTATCCGAGAGCTGCTGGACCGCTGTACAGGAAGGCGTGGTGAC 878

Qy 181 GGAGCCGAGTCAAGTCTACATCCAGCAGCTGGTGGAGGGCTGCACTACCTGCAAG 240

Db 879 GGAGCCGAGTCAAGTCTACATCCAGCAGCTGGTGGAGGGCTGCACTACCTGCAAG 938

Qy 241 CCATGGCTTCTCCACCTGGACATAAAGCCCTCTAAACATCTGATGGTGCATCTGCCCG 300

Db 939 CCATGGCTTCTCCACCTGGACATAAAGCCCTCTAAACATCTGATGGTGCATCTGCCCG 998

Qy 301 GGAAGACATTAAATCTGCGACTTTGGCTTTGGCCAGAACATACCCAGCAGAGCTGCA 360

Db 999 GGAAGACATTAAATCTGCGACTTTGGCTTTGGCCAGAACATACCCAGCAGAGCTGCA 1058

Qy 361 GTTCAGCCAGTACGGCTCCCTGATGGTCTCTCCCGAGATCATCCAGCAGAACCTGT 420

Db 1059 GTTCAGCCAGTACGGCTCCCTGATGGTCTCTCCCGAGATCATCCAGCAGAACCTGT 1118

Qy 421 GAGCGAAGCTCCGACATTTGGGCGCATGGGTGTCTCTCTACCTCAGCCCTGACCTGCTC 480

Db 1119 GAGCGAAGCTCCGACATTTGGGCGCATGGGTGTCTCTCTACCTCAGCCCTGACCTGCTC 1178

Qy 481 ATCCCCATTTGCCGGCGAGAGTGACCGGTGCCACCTCTCTGAAACCTCTTGGAGGGCGGT 540

Db 1179 ATCCCCATTTGCCGGCGAGAGTGACCGGTGCCACCTCTCTGAAACCTCTTGGAGGGCGGT 1238

Qy 541 GTATGAGCAGCCCCATGGTGGCCCTCCTCAGCGAAGACGCCAAAGACTTTCATCAAGGC 600

Db 1239 GTATGAGCAGCCCCATGGTGGCCCTCCTCAGCGAAGACGCCAAAGACTTTCATCAAGGC 1298

Qy 601 TACGCTCAGAGACCCCTCAGGCGCGGCTAGTGGCGCCAGTGCCTCTCCACCCCTG 660

Db 1299 TACGCTCAGAGACCCCTCAGGCGCGGCTAGTGGCGCCAGTGCCTCTCTCCACCCCTG 1358

Qy 661 GTTCCTGAAATTCATGCTCGGAGGAGGCCACTTCATCAACCAAGCAGCTCAAGTT 720

Db 1359 GTTCCTGAAATTCATGCTCGGAGGAGGCCACTTCATCAACCAAGCAGCTCAAGTT 1418

Qy 721 CCTCTGCGCCGAAAGTCTGCGCAGCGTTCCTCATGAGCTACAGTCCATCTCTGGTGAT 780

Db 1419 CCTCTGCGCCGAAAGTCTGCGCAGCGTTCCTCATGAGCTACAGTCCATCTCTGGTGAT 1478

Qy 781 GCGCTCCATCTCCTGAGCTGTGCGGGGCCACCCGACAGACCCCTCTCTCGCGCTAGCCCG 840

Db 1479 GCGCTCCATCTCCTGAGCTGTGCGGGGCCACCCGACAGACCCCTCTCTCGCGCTAGCCCG 1538

Qy 841 GCACCTCTGACGAGGACACTGGTGGCTCTCTCAGTTCTCTCTCTCTCTCTGACACAGCT 900

Db 1539 GCACCTCTGACGAGGACACTGGTGGCTCTCTCAGTTCTCTCTCTCTCTCTGACACAGCT 1598

Qy 901 CGCCCATTTTCCCGGGCTAAGTCACTGCGACCCCTCCCGGTGACACACTCACCACCTGCT 960

Db 1599 CGCCCATTTTCCCGGGCTAAGTCACTGCGACCCCTCCCGGTGACACACTCACCACCTGCT 1658

Qy 961 GCACCCCGGGGCTTCTGCGGGCTCTCGGCGCTCTGAGGAGCCGAGGCGCAGTGA 1020

Db 1659 GCACCCCGGGGCTTCTGCGGGCTCTCGGCGCTCTGAGGAGCCGAGGCGCAGTGA 1718

Qy 1021 GCGCTCCACGAGGCCCTCAGCTCGGCTGCTCTCCGAGGGTCTCGGGCCACCGCGCGC 1080

Db 1719 GCGCTCCACGAGGCCCTCAGCTCGGCTGCTCTCCGAGGGTCTCGGGCCACCGCGCGC 1778

Qy 1081 CAGGGCTGCTGCGCCCGGACAGCTCATCCGAGCTCTTCTACCAACAGGCGGGTGA 1140

Db 1779 CAGGGCTGCTGCGCCCGGACAGCTCATCCGAGCTCTTCTACCAACAGGCGGGTGA 1838

Qy 1141 GAGCCCTGAGCAGCGGGCTCTGCGCGGAGCAGCGCGCCCGCGCGCGCGCA 1200

Db 1839 GAGCCCTGAGCAGCGGGCTCTGCGCGGAGCAGCGCGCCCGCGCGCGCGCA 1898

Qy 1201 CTGCTGAAAGCGGGTACATTTGCGGGGGCTGCGAGGCTGCGCGAGCCACTGATGGA 1260

Db 1899 CTGCTGAAAGCGGGTACATTTGCGGGGGCTGCGAGGCTGCGCGAGCCACTGATGGA 1958

Qy 1261 GCACCGGCTCTGAGGAGGAGCGCGCAGGAGGAGCAGCCACCTCTCTGGGCCAAAGC 1320

Db 1959 GCACCGGCTCTGAGGAGGAGCGCGCAGGAGGAGCAGCCACCTCTCTGGGCCAAAGC 2018

Qy 1321 CCCCTCATTTGAGACTGCCCTCGGCTGCTCTGCGACCCCACTTTGGCCCCCTGGCGCA 1380

Db 2019 CCCCTCATTTGAGACTGCCCTCGGCTGCTCTGCGACCCCACTTTGGCCCCCTGGCGCA 2078

Qy 1381 CAGCCACTCCCTGGAACATGACTCTCGAGCAGCCCCCGCGCCCTCTCTGAGGCGCTGCGG 1440

Db 2079 CAGCCACTCCCTGGAACATGACTCTCGAGCAGCCCCCGCGCCCTCTCTGAGGCGCTGCGG 2138

Qy 1441 TGAGGCAACGACTGCTTTCAGCCCTCTCGGGGGGGCCCTATCAGGAGCATGGGGCA 1500

Db 2139 TGAGGCAACGACTGCTTTCAGCCCTCTCGGGGGGGGGCCCTATCAGGAGCATGGGGCA 2198

Qy 1501 CCCTCAGGGCTCCAAAGCAGCTTCCATCCACTGTGGCCACCCAGGCACTGTCTAGCCAGA 1560

Db 2199 CCCTCAGGGCTCCAAAGCAGCTTCCATCCACTGTGGCCACCCAGGCACTGTCTAGCCAGA 2258

Qy 1561 GAGGCCATCCCGGACAGCCCTTGGGGGCGAGCGCCCTTTTGTGCAACCCCAAGCAGGG 1620

Db 2259 GAGGCCATCCCGGACAGCCCTTGGGGGCGAGCGCCCTTTTGTGCAACCCCAAGCAGGG 2318

Qy 1621 TTCTGCCCCCGAGGAGGCTGACGCCCCCAAGCAGTTCGCCCATGCTCTCTGGCTC 1680

Db 2319 TTCTGCCCCCGAGGAGGCTGACGCCCCCAAGCAGTTCGCCCATGCTCTCTGGCTC 2378

Qy 1681 CTTCCCTCCAGGATCTTGGCAAGAGGCGCCCTTAGTACCTCTCAAGCCCTCTCTTGGGACA 1740

Db 2379 CTTCCCTCCAGGATCTTGGCAAGAGGCGCCCTTAGTACCTCTCAAGCCCTCTCTTGGGACA 2438

QY 1741 GCCCAGGACCCCTGCCCCCTGCCAAGCAAGACGCCCCCAATTGGACTCTAAGATGGGCC 1800
Db 2439 GCCCAGGACCCCTGCCCCCTGCCAAGCAAGACGCCCCCAATTGGACTCTAAGATGGGCC 2498
QY 1801 TGGAGACATCTCTCTTCTGGAGAGCAAAACCCGGCCCTGACGTCTCCAGAGGTGAGC 1860
Db 2499 TGGAGACATCTCTCTTCTGGAGAGCAAAACCCGGCCCTGACGTCTCCAGAGGTGAGC 2558
QY 1861 CTCCAGGCGAGCTCTTCCCAAGTAGCTCTCCCTCAGGGTGGGCTCTCTCCAGAGTGGGCAC 1920
Db 2559 CTCCAGGCGAGCTCTTCCCAAGTAGCTCTCCCTCAGGGTGGGCTCTCTCCAGAGTGGGCAC 2618
QY 1921 AGAGCTGGCCCTCTCCCTGATGCGAGGGCTGGAACCCAGAGGCTGAGGATCTGTCCGA 1980
Db 2619 AGAGCTGGCCCTCTCCCTGATGCGAGGGCTGGAACCCAGAGGCTGAGGATCTGTCCGA 2678
QY 1981 CTCCACACCCACTTGCAGGCGCTCAGGAACAGSCGACCATGCGCAAGTTCTCCCTGGG 2040
Db 2679 CTCCACACCCACTTGCAGGCGCTCAGGAACAGGTGACCATGCGCAAGTTCTCCCTGGG 2738
QY 2041 TGGTCGCGGGGCTACCGAGGCGTGGCTGCTATGGCACCTTTGGCTTTGGTGAGATGC 2100
Db 2739 TGGTCGCGGGGCTACCGAGGCGTGGCTGCTATGGCACCTTTGGCTTTGGTGAGATGC 2798
QY 2101 AGGGGCACTGTTGGGCAAGGGCCCATGTGGCCAGGATAGCTGGGCTGTGTCCAGTC 2160
Db 2799 AGGGGCACTGTTGGGCAAGGGCCCATGTGGGCGAGGATAGCTGGGCTGTGTCCAGTC 2858
QY 2161 GGAGGAGGAGCAGAGGAGGCGCAGGCTTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2220
Db 2859 GGAGGAGGAGCAGAGGAGGCGCAGGCTTGAGTCCAGTCCGAGGAGCAGCAGGAGGC 2918
QY 2221 CAGGCTGAGAGCCACTGCCCCAGGTGAGTGCAGAGGCTGTGCTGAGTTCGCGAGGC 2280
Db 2919 CAGGCTGAGAGCCACTGCCCCAGGTGAGTGCAGAGGCTGTGCTGAGTTCGCGAGGC 2978
QY 2281 TCCACACGAGGCTCTCCAGAGCCACCCCATGCGAGGAGACATCGGGAGGTCTCCCTGGT 2340
Db 2979 TCCACACGAGGCTCTCCAGAGCCACCCCATGCGAGGAGACATCGGGAGGTCTCCCTGGT 3038
QY 2341 GCAGATCGGGACCTGTGAGTGATGGGGAGGCGCGCACAAATATCCCTGGACATTC 2400
Db 3039 GCAGATCGGGACCTGTGAGTGATGGGGAGGCGCGCACAAATATCCCTGGACATTC 3098
QY 2401 CGAGGTGAGCCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db 3099 CGAGGTGAGCCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 3158
QY 2461 CGAGTTTATGATCTTCAGGAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCCAT 2520
Db 3159 CGAGTTTATGATCTTCAGGAAGTCCCAAGTCCGCTCAGCCAGAGCGCCCTCCCCCAT 3218
QY 2521 GGCTGAGGAGAGTGGCCGAGTTCCCGAGGCCACGCTGGCCCTTGGCCAGGTGAATGGG 2580
Db 3219 GGCTGAGGAGAGTGGCCGAGTTCCCGAGGCCACGCTGGCCCTTGGCCAGGTGAATGGG 3278
QY 2581 CCCCCACGAGGCTGAGATCAAGAGAGTCAAGAGATGTGACGCGCTGTGCGAGA 2640
Db 3279 CCCCCACGAGGCTGAGATCAAGAGAGTCAAGAGATGTGACGCGCTGTGCGAGA 3338
QY 2641 GGCTGCGGTGGGAGGAAGCGCAAGTGTCTCGCGCTCAGCGAGCTCTTCCACTTCCC 2700
Db 3339 GGCTGCGGTGGGAGGAAGCGCAAGTGTCTCGCGCTCAGCGAGCTCTTCCACTTCCC 3398
QY 2701 TGGAGGCACTGCGCGTGTAGCTGAGAGCTGGGCTGCGTGAGAGTGAAGGC 2760
Db 3399 TGGAGGCACTGCGCGTGTAGCTGAGAGCTGGGCTGCGTGAGAGTGAAGGC 3458
QY 2761 CTCGTTGAGACATCTCCCGGATCTTGAAGGCGAGCCGGAGGTCTGGAAGAGGAGG 2820
Db 3459 CTCGTTGAGACATCTCCCGGATCTTGAAGGCGAGCCGGAGGTCTGGAAGAGGAGG 3518
QY 2821 GCCCCCCAGGAAGACCGCCCTTGCTTCTCCGCTCTCAGGTCTCTGAAGAGCTGGGA 2880

Db 3519 GCCCCCCAGGAAGCCAGGCTTGTCTTCTCCGCTCTCAGGTCTGAAGAGCTGGGA 3578
QY 2881 CCAGAGCCCGACATCTCTTAAGGAGCTCTCAGATGAGACTGTGTCCTGGGCCAGTCAGT 2940
Db 3579 CCAGAGCCCGACATCTCTTAAGGAGCTCTCAGATGAGACTGTGTCCTGGGCCAGTCAGT 3638
QY 2941 GACACTGGCTGCGCAGGTGTGAGCCAGCAGCTGCCAGGCGACCTGGAGCAAGACGG 3000
Db 3639 GACACTGGCTGCGCAGGTGTGAGCCAGCAGCTGCCAGGCGACCTGGAGCAAGACGG 3698
QY 3001 A 3001
Db 3699 A 3699

RESULT 9
US-10-274-978-1
; Sequence 1, Application US/10274978
; Publication No. US20030064475A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274, 978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
US-10-274-978-1

Query Match 99.9%; Score 2997.8; DB 14; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCTACGGAGCAGAACTCGGGCCCGAGGCATACAGGAGCGAGACATCCTGGCCGCGCTGAG 60
Db 699 CCTACGGAGCAGAACTCGGGCCCGAGGCATACAGGAGCGAGACATCCTGGCCGCGCTGAG 758
QY 61 CCACCCGCTGGTCA CGGGCTGTGGA CCAGTTTGAGACCCGCAAGACCTCATCTCAT 120
Db 759 CCACCCGCTGGTCA CGGGCTGTGGA CCAGTTTGAGACCCGCAAGACCTCATCTCAT 818
QY 121 CTTGGAGCTGTCTCATCCGAGAGCTGTGACCCGCTGTACAGGAAGGGCTGTGTGAC 180
Db 819 CTTGGAGCTGTCTCATCCGAGAGCTGTGACCCGCTGTACAGGAAGGGCTGTGTGAC 878
QY 181 GGAGGCCGAGGTCAAGGTCTACATCCAGCAGCTGGTGAGGGGCTGCATCTACCTGCACAG 240
Db 879 GGAGGCCGAGGTCAAGGTCTACATCCAGCAGCTGGTGAGGGGCTGCATCTACCTGCACAG 938
QY 241 CCATGGGCTTCTCCACCTGGACATTAAGCCCTCTAAATCTCTGATGGTGATCTTCGCCG 300
Db 939 CCATGGGCTTCTCCACCTGGACATTAAGCCCTCTAAATCTCTGATGGTGATCTTCGCCG 998
QY 301 GGAAGACATTAAATCTCGGACTTTGGCTTTGCCCGAGAACATACCCCGAGAGAGCTGCA 360
Db 999 GGAAGACATTAAATCTCGGACTTTGGCTTTGCCCGAGAACATACCCCGAGAGAGCTGCA 1058
QY 361 GTTACGACGATACGGCTCCCTCCTGAGTGTCTTCCCGGAGATCATCCAGAGAACCTCTGT 420
Db 1059 GTTACGACGATACGGCTCCCTCCTGAGTGTCTTCCCGGAGATCATCCAGAGAACCTCTGT 1118
QY 421 GAGCGAAGCCTCCGACATTTGGGCCATTTGGGGCTGTCTATCTTACCTACCTGACCTGCTC 480


```
QY 2641 GGCTGCGTGGGAGGAGCGCAAGTGGTCTCGCGGTCAAGAGCGCTCTTCACTTCCC 2700
Db 3339 GGCTGCGTGGGAGGAGCGCAAGTGGTCTCGCGGTCAAGAGCGCTCTTCACTTCCC 3398
QY 2701 TGGGAGGCACTTGGCGCTGGATGAGCTGCAAGTGGGGTGGCTGAGAGAGTGAAGGC 2760
Db 3399 TGGGAGGCACTTGGCGCTGGATGAGCTGCAAGTGGGGTGGCTGAGAGAGTGAAGGC 3458
QY 2761 CTCCGTGGAGCACATCTCCGGATCCTGAAGGCGAGCGGAGAGTCTGGAGAGGAGGG 2820
Db 3459 CTCCGTGGAGCACATCTCCGGATCCTGAAGGCGAGCGGAGAGTCTGGAGAGGAGGG 3518
QY 2821 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2880
Db 3519 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3578
QY 2881 CCGAGCGGCGACATCTTAAGGAGCTCTCAGATGAGACTGTGTGCTTGGGCGCAGTCA 2940
Db 3579 CCGAGCGGCGACATCTTAAGGAGCTCTCAGATGAGACTGTGTGCTTGGGCGCAGTCA 3638
QY 2941 GACACTGGCTGCCAGGTGTAGCCAGCGAGCTGCCAGGCGCAGCTGGAGCAAGAGCG 3000
Db 3639 GACACTGGCTGCCAGGTGTAGCCAGCGAGCTGCCAGGCGCAGCTGGAGCAAGAGCG 3698
QY 3001 A 3001
Db 3699 A 3699

RESULT 10
US-10-274-978-3
; Sequence 3, Application US/10274978
; Publication No. US20030064475A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
; US-10-274-978-3

Query Match 99.9%; Score 2997.8; DB 14; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCTACGGAGCAGAACTCGGGCCAGGCATACAGGGAGCGAGACATCTCTGGCGCGCTGAG 60
Db 699 CCTACGGAGCAGAACTCGGGCCAGGCATACAGGGAGCGAGACATCTCTGGCGCGCTGAG 758
QY 61 CCACCCGCTGGTCAAGGCTGTGGACCAAGTTTGAAGCCGCAAGACCTCATCTCAT 120
Db 759 CCACCCGCTGGTCAAGGCTGTGGACCAAGTTTGAAGCCGCAAGACCTCATCTCAT 818
QY 121 CTTGAGCTGTGTATCCAGGAGCTGTGGACCGCTGTACAGAAAGGCGTGTGAC 180
Db 819 CTTGAGCTGTGTATCCAGGAGCTGTGGACCGCTGTACAGAAAGGCGTGTGAC 878
QY 181 GGAGGCGAGGTCAAGGTCTATCCAGCAGCTGTGGAGGGGCTGCATCTACCTGCACAG 240
Db 879 GGAGGCGAGGTCAAGGTCTATCCAGCAGCTGTGGAGGGGCTGCATCTACCTGCACAG 938
```

```
QY 241 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTAAACATCTCTGATGTGATCTGCGCG 300
Db 939 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTAAACATCTCTGATGTGATCTGCGCG 998
QY 301 GGAAGACATTTAAATCTGCGACTTTGGCTTTGGCCAGAAATCACCCACAGAGCTGCA 360
Db 999 GGAAGACATTTAAATCTGCGACTTTGGCTTTGGCCAGAAATCACCCACAGAGCTGCA 1058
QY 361 GTTCAGCAGTACGGCTCCCTGAGTTCTCTCCCGAGATCATCCAGAGAACCTGT 420
Db 1059 GTTCAGCAGTACGGCTCCCTGAGTTCTCTCCCGAGATCATCCAGAGAACCTGT 1118
QY 421 GAGGAGAGCTCCGACATTTGGGCGCATGGGTGTCATCTCTACTCAGCCTGACCTGCTC 480
Db 1119 GAGGAGAGCTCCGACATTTGGGCGCATGGGTGTCATCTCTACTCAGCCTGACCTGCTC 1178
QY 481 ATCCCATTTGCGGGGAGAGTGACCGTGGCACCTCTCTGAAACGTCTCTGGAGGGGCGCT 540
Db 1179 ATCCCATTTGCGGGGAGAGTGACCGTGGCACCTCTCTGAAACGTCTCTGGAGGGGCGCT 1238
QY 541 GTTCATGAGCAGCCCCATGGCTGCCCATCTCAGCGAAGACGCCAAAGACTTCATCAAGGC 600
Db 1239 GTTCATGAGCAGCCCCATGGCTGCCCATCTCAGCGAAGACGCCAAAGACTTCATCAAGGC 1298
QY 601 TAGCGTGCAGAGAGCCCTCAGGSCCGGCTAGTGGCGGCCAGTGCTCTCCACCCCTG 660
Db 1299 TAGCGTGCAGAGAGCCCTCAGGSCCGGCTAGTGGCGGCCAGTGCTCTCCACCCCTG 1358
QY 661 GTTCTGAAATCCATGCTCGGAGAGGCGCCACTTTCATCAACACCAAGCAGCTCAAGTT 720
Db 1359 GTTCTGAAATCCATGCTCGGAGAGGCGCCACTTTCATCAACACCAAGCAGCTCAAGTT 1418
QY 721 CTTCTGGGCGGAGTGGCTGGCAGAGTTCCCTGATGAGCTCAAGTCCATCTCTGTGAT 780
Db 1419 CTTCTGGGCGGAGTGGCTGGCAGAGTTCCCTGATGAGCTCAAGTCCATCTCTGTGAT 1478
QY 781 GCGCTCCATCTCAGCTGCTCGGGGCGCCACCCGACAGCCCTCTCTGGGGTAGCCCCG 840
Db 1479 GCGCTCCATCTCAGCTGCTCGGGGCGCCACCCGACAGCCCTCTCTGGGGTAGCCCCG 1538
QY 841 GCACCTCTGAGGAGACATGTTGGTCTCTCCAGTTCTCTCTCTCTGACAAAGAGCT 900
Db 1539 GCACCTCTGAGGAGACATGTTGGTCTCTCCAGTTCTCTCTCTCTGACAAAGAGCT 1598
QY 901 CGCCCATTTGGCGGCTTAAGTCACTGCGACCTCTCCCGGTGACACACTCACCAGCT 960
Db 1599 CGCCCATTTGGCGGCTTAAGTCACTGCGACCTCTCCCGGTGACACACTCACCAGCT 1658
QY 961 GCACCCCGGGGCTTCTGCGGCGCTCGGCGCAGCTGCTGAGGAGCGAGCGCAGTGA 1020
Db 1659 GCACCCCGGGGCTTCTGCGGCGCTCGGCGCAGCTGCTGAGGAGCGAGCGCAGTGA 1718
QY 1021 GCGCTCCACCGAGGCGCCAGTCTCGCTGCACTCTCCGAGGTGCGGCGCCACCGCGCGC 1080
Db 1719 GCGCTCCACCGAGGCGCCAGTCTCGCTGCACTCTCCGAGGTGCGGCGCCACCGCGCGC 1778
QY 1081 CCAGGCGTGGTGGCGGCGCAGAGTCACTCGGAGCTGTTTCTACCAAGCGGCTGA 1140
Db 1779 CCAGGCGTGGTGGCGGCGCAGAGTCACTCGGAGCTGTTTCTACCAAGCGGCTGA 1838
QY 1141 GAGCCCTGAGCAGCGGGGCGCTGGGCGCGGAGCAGCGGCGCACCCGCGCGCGCGCA 1200
Db 1839 GAGCCCTGAGCAGCGGGGCGCTGGGCGCGGAGCAGCGGCGCACCCGCGCGCGCGCA 1898
QY 1201 CTTGCTGAAGGGCGGTACATTGGGGGCGCTGCCAGGCTGCGCGAGGACCTGATGA 1260
Db 1899 CTTGCTGAAGGGCGGTACATTGGGGGCGCTGCCAGGCTGCGCGAGGACCTGATGA 1958
QY 1261 GCACCGCTGTGGAGGAGGAGGCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
Db 1959 GCACCGCTGTGGAGGAGGAGGCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2018
```

1321 CCCCTCATTGAGACTGCCCTCCGGCTGCCCTTGGGACCCACTTTGGGCCCTTGGCA 1380
Db
2019 CCCCTCATTCAGACTGCCCTCCGGCTGCCCTTGGGACCCACTTTGGGCCCTTGGCA 2078
Qy
1381 CAGCACTCCCTGGAACATGACTCTCCGAGACACCCCGCCCTCTCCGAGGCTTGGG 1440
Db
2079 CAGCACTCCCTGGAACATGACTCTCCGAGACACCCCGCCCTCTCCGAGGCTTGGG 2138
Qy
1441 TGAGGCAAGCACTGCTTTCAGCCCTTCCGCGGGGGCCCTATCAAGGACATGGGGCA 1500
Db
2139 TGAGGCAAGCACTGCTTTCAGCCCTTCCGCGGGGGCCCTATCAAGGACATGGGGCA 2198
Qy
1501 CCCTCAGGCTCCCAAGCAGCTTCCATCCACTGGTGGGACACCCAGGCACTGTCAGCCAGA 1560
Db
2199 CCCTCAGGCTCCCAAGCAGCTTCCATCCACTGGTGGGACACCCAGGCACTGTCAGCCAGA 2258
Qy
1561 GAGGCCATCCCGGACAGCCCTTGGGGGACCCAGCCCTTTCGCCACCCCAAGCAGGG 1620
Db
2259 GAGGCCATCCCGGACAGCCCTTGGGGGACCCAGCCCTTTCGCCACCCCAAGCAGGG 2318
Qy
1621 TTCTGCCCCCAGAGGGCTGAGCCCCCAACCCAGCAGTTGCCCATGCCCTCTCTGGCTC 1680
Db
2319 TTCTGCCCCCAGAGGGCTGAGCCCCCAACCCAGCAGTTGCCCATGCCCTCTCTGGCTC 2378
Qy
1681 CTTCCTCCAGGATCTTGAAGAGGCCCCCTTAGTACCTTCAAGCCCCCTTTTGGGACA 1740
Db
2379 CTTCCTCCAGGATCTTGAAGAGGCCCCCTTAGTACCTTCAAGCCCCCTTTTGGGACA 2438
Qy
1741 GCCCCAGGACCCCTGCCCCGCAAGCAAGCCCCCCTTGGACTTGAAGTGGGGCC 1800
Db
2439 GCCCCAGGACCCCTGCCCCGCAAGCAAGCCCCCCTTGGACTTGAAGTGGGGCC 2498
Qy
1801 TGGAGACATCTCTTCTCTGGAGGCCCCAAACCCCGCCCCCTTGCAGTTTCCCGAGGTCAGC 1860
Db
2499 TGGAGACATCTCTTCTCTGGAGGCCCCAAACCCCGCCCCCTTGCAGTTTCCCGAGGTCAGC 2558
Qy
1861 CTCCAGGCGAGCTTTCAGAGTGAAGTCCCTCAGGTGGGCTCCTCCAGGTGGGCAC 1920
Db
2559 CTCCAGGCGAGCTTTCAGAGTGAAGTCCCTCAGGTGGGCTCCTCCAGGTGGGCAC 2618
Qy
1921 AGAGCTGGCCCCCTCCCTGGATGGAGGGCTGACCCAGAGGCTGAGGATCTGTCCGA 1980
Db
2619 AGAGCTGGCCCCCTCCCTGGATGGAGGGCTGACCCAGAGGCTGAGGATCTGTCCGA 2678
Qy
1981 CTCACACCCACCTTGACGCGCCCTCAGGAACAGCGCACCATGCGCAAGTTTCTCCCTGGG 2040
Db
2679 CTCACACCCACCTTGACGCGCCCTCAGGAACAGCGCACCATGCGCAAGTTTCTCCCTGGG 2738
Qy
2041 TGGTCCGGGGGCTACGAGCGGTGGCTGGCTATGGCACTTTGGCTTGGTGGAGATGC 2100
Db
2739 TGGTCCGGGGGCTACGAGCGGTGGCTGGCTATGGCACTTTGGCTTGGTGGAGATGC 2798
Qy
2101 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCGAGGATAGCTGGGCTGTGTCCAGTC 2160
Db
2799 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCGAGGATAGCTGGGCTGTGTCCAGTC 2858
Qy
2161 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2220
Db
2859 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2918
Qy
2221 CAGGCTGAGAGCCCATGTCGCCCGAGGTGAGTGCAGAGGCTGTGCTGAGGTTCGGCAGGGC 2280
Db
2919 CAGGCTGAGAGCCCATGTCGCCCGAGGTGAGTGCAGAGGCTGTGCTGAGGTTCGGCAGGGC 2978
Qy
2281 TCCACACAGGAGCTTCCAGAGCCACCCCATGGAGGACATCGGGCAGGTCTCCCTGGT 2340
Db
2979 TCCACACAGGAGCTTCCAGAGCCACCCCATGGAGGACATCGGGCAGGTCTCCCTGGT 3038
Qy
2341 GCAGATCCGGACCTGTGAGGTGATGGGAGGCGGCGACACAATATCCCTGGACATTTTC 2400
Db
3039 GCAGATCCGGACCTGTGAGGTGATGGGAGGCGGCGACACAATATCCCTGGACATTTTC 3098
Qy
2401 CGAGGTGACCCCGCTTACCTCAACCTCTCAGACCTCTGACATATCAAGTACCTCCCATTT 2460

Db
3099 CGAGGTGAGACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCATTT 3158
Qy
2461 CGAGTTTATGATCTTTCAGGAAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 2520
Db
3159 CGAGTTTATGATCTTTCAGGAAAGTCCCAAGTCCGCTCAGCAGAGCGCCCTCCCCCAT 3218
Qy
2521 GGCTGAGGAGAGCTGGCCGAGTTTCCGAGGCCACAGTGGCCCTTGGCCAGGTGAATGGG 2580
Db
3219 GGCTGAGGAGAGCTGGCCGAGTTTCCGAGGCCACAGTGGCCCTTGGCCAGGTGAATGGG 3278
Qy
2581 CCCCCAGCAGGCTGGAGATCACAGAGGAGTCAAGAGTGTGACGCGCTGTGGCAGA 2640
Db
3279 CCCCCAGCAGGCTGGAGATCACAGAGGAGTCAAGAGTGTGACGCGCTGTGGCAGA 3338
Qy
2641 GGCTGCCGTGGGCAAGGCGCAAGTGTCTTCCGCTCAGCAGCCTTTCCTCACTTCCC 2700
Db
3339 GGCTGCCGTGGGCAAGGCGCAAGTGTCTTCCGCTCAGCAGCCTTTCCTCACTTCCC 3398
Qy
2701 TGGAGGACCTTCCGCTGGATGAGCTTGCAGAGCTGGGCTGCTGAGAGTGAAGGC 2760
Db
3399 TGGAGGACCTTCCGCTGGATGAGCTTGCAGAGCTGGGCTGCTGAGAGTGAAGGC 3458
Qy
2761 CTCGTCGAGCACATCTCCCGATCTTAAAGGGCAGGCCGAGAGTCTGGAGAAAGAGGG 2820
Db
3459 CTCGTCGAGCACATCTCCCGATCTTAAAGGGCAGGCCGAGAGTCTGGAGAAAGAGGG 3518
Qy
2821 GCCCCCAGGAAGAGCAGGCTTGTCTTCTCCGCTCTCAGGTCTTGAAGAGCTGGGA 2880
Db
3519 GCCCCCAGGAAGAGCAGGCTTGTCTTCTCCGCTCTCAGGTCTTGAAGAGCTGGGA 3578
Qy
2881 CCGAGCGCCGACATTCCTAAGGGAGCTCTCAGATGAGACTGTGTCTTGGGCCAGTCACT 2940
Db
3579 CCGAGCGCCGACATTCCTAAGGGAGCTCTCAGATGAGACTGTGTCTTGGGCCAGTCACT 3638
Qy
2941 GACACTGCGCTGCCAGGTGTGAGCCAGCAGCTGCCAGCCACCTTGGAGCAAGAGCGG 3000
Db
3639 GACACTGCGCTGCCAGGTGTGAGCCAGCAGCTGCCAGCCACCTTGGAGCAAGAGCGG 3698
Qy
3001 A 3001
Db
3699 A 3699

RESULT 11

US-10-697-263-1
; Sequence 1, Application US/10697263
; Publication No. US20040063142A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00927-CIP-DIV2
; CURRENT APPLICATION NUMBER: US/10/697,263
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-697-263-1

Query Match 99.9%; Score 2997.8; DB 17; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTACGAGCAGAACTCGGGCCAGGCATACAGGAGCGAGACATCCTTGCCGCGCTGAG 60
Db |||||
699 CTTACGAGCAGAACTCGGGCCAGGCATACAGGAGCGAGACATCCTTGCCGCGCTGAG 758
QY 61 CCACCCGCTGCTCACGGGCTGCTGACACAGTTTGAGACCCGAGAACCTCATCTCAT 120
Db |||||
759 CCACCCGCTGCTCACGGGCTGCTGACACAGTTTGAGACCCGAGAACCTCATCTCAT 818
QY 121 CTTGAGCTGTGCTCATCCGAGAGCTGTGGACCGCTGTACAGAGAGGGCTGGTGAC 180
Db |||||
819 CTTGAGCTGTGCTCATCCGAGAGCTGTGGACCGCTGTACAGAGAGGGCTGGTGAC 878
QY 181 GAGGCCGAGGCTCAAGGTTCTATCATCAGCAGCTGTGGAGGGCTGCACTACTCTGCACAG 240
Db |||||
879 GGAGGCCGAGGCTCAAGGTTCTATCATCAGCAGCTGTGGAGGGCTGCACTACTCTGCACAG 938
QY 241 CCATGCGCTTCCACCTCGGACATTAAGCCCTCTAACCATCTCTGATGCTGCTGCCG 300
Db |||||
939 CCATGCGCTTCCACCTCGGACATTAAGCCCTCTAACCATCTCTGATGCTGCTGCCG 998
QY 301 GGAAGACATTAATACTCGGACTTTGGCTTTTGGCCAGAACATCACCCAGCAGAGCTGCA 360
Db |||||
999 GGAAGACATTAATACTCGGACTTTGGCTTTGCCCCAGAACATCACCCAGCAGAGCTGCA 1058
QY 361 GTTACGCCAGTACGGCTCCCTGAGTTCGTCTCCCGAGATCATCCAGCAGAACCCCTGT 420
Db |||||
1059 GTTACGCCAGTACGGCTCCCTGAGTTCGTCTCCCGAGATCATCCAGCAGAACCCCTGT 1118
QY 421 GAGCGAAGCTCCGACATTTGGCCATGSGTGTATCTCTCTACTCTGATGCTGCTC 480
Db |||||
1119 GAGCGAAGCTCCGACATTTGGCCATGSGTGTATCTCTCTACTCTGATGCTGCTC 1178
QY 481 ATCCCAATTTGCCGCGAGAGTGACCGTGCACACCTCTCTGAACGTCTCTGAGGGCGCGT 540
Db |||||
1179 ATCCCAATTTGCCGCGAGAGTGACCGTGCACACCTCTCTGAACGTCTCTGAGGGCGCGT 1238
QY 541 GTCATGAGACGCCCCATGCTGCCACCTCAGCGAAGACGCCAAGACTTCAACAAGGC 600
Db |||||
1239 GTCATGAGACGCCCCATGCTGCCACCTCAGCGAAGACGCCAAGACTTCAACAAGGC 1298
QY 601 TACGCTCAGAGAGCCCCCTCAGGCCCGGCTAGTGCGCCCGCAGTGCTCTCCACCCCTG 660
Db |||||
1299 TACGCTCAGAGAGCCCCCTCAGGCCCGGCTAGTGCGCCCGCAGTGCTCTCCACCCCTG 1358
QY 661 GTTCTGAAATCCATGCTCGGAGAGGCCCCATTATCAACAACCAAGCAGCTCAAGTT 720
Db |||||
1359 GTTCTGAAATCCATGCTCGGAGAGGCCCCATTATCAACAACCAAGCAGCTCAAGTT 1418
QY 721 CCTCTGSCCGAAGTCTGCTGCGAGCGTTCCCTGATGAGCTACAAAGTCCATCTCTGAT 780
Db |||||
1419 CCTCTGSCCGAAGTCTGCTGCGAGCGTTCCCTGATGAGCTACAAAGTCCATCTCTGAT 1478
QY 781 GCGTCTCATCCCTGAGCTGTGCGGGCCACACCCGACAGCCCTCTCCCTCGGCGTAGCCCG 840
Db |||||
1479 GCGTCTCATCCCTGAGCTGTGCGGGCCACACCCGACAGCCCTCTCCCTCGGCGTAGCCCG 1538
QY 841 GCACCTCTGAGGACACTGTGTGCTCCTCAGTTCTCTCTCTCTGACAAACAGCT 900
Db |||||
1539 GCACCTCTGAGGACACTGTGTGCTCCTCAGTTCTCTCTCTCTGACAAACAGCT 1598
QY 901 CGCCCAATTTGCCGGGCTAAGTCACTGCCACCTCTCCCGTGACACACTCACTGCT 960
Db |||||
1599 CGCCCAATTTGCCGGGCTAAGTCACTGCCACCTCTCCCGTGACACACTCACTGCT 1658
QY 961 GCACCCCGGGCTTCTGCGGCCCTGCGCAGCCTGCTTGAGGAAGCCGAGGCCAGTGA 1020
Db |||||
1659 GCACCCCGGGGTTCTGCGGCCCTCGGCAGCCTGCTTGAGGAAGCCGAGGCCAGTGA 1718
QY 1021 GCGTCTCACCGAGCCCGCTCGCTGATCTCCGAGGGTGTCCGGGCCACCGGCCG 1080
Db |||||
1719 GCGTCTCACCGAGCCCGCTCGCTGATCTCCGAGGGTGTCCGGGCCACCGGCCG 1778

QY 1081 CCAGGGCTGCTGCTGCCCGGACACAGGCTATCCGCGAGGCTGTTCTACCAACAGGCGGGTGA 1140
Db |||||
1779 CCAGGGCTGCTGCTGCCCGGACACAGGCTATCCGCGAGGCTGTTCTACCAACAGGCGGGTGA 1838
QY 1141 GAGCCCTTGAGCAGCGGGCCCTTGCGCCCGGGGAGCAGGCGGCACCCGCGCCGCGCGCGCA 1200
Db |||||
1839 GAGCCCTTGAGCAGCGGGCCCTTGCGCCCGGGGAGCAGGCGGCACCCGCGCCGCGCGCGCA 1898
QY 1201 CTTGCTGAAGGGCGGGTACATTTGGGGGGGCTGTGCGAGGCTGTGCGAGCACTGATGGA 1260
Db |||||
1899 CTTGCTGAAGGGCGGGTACATTTGGGGGGGCTGTGCGAGGCTGTGCGAGCACTGATGGA 1958
QY 1261 GCACCGGCTGTGAGGAGGAGCGCGCAGGAGGAGCAGGCGCACCCCTCTCTGCGCAAGC 1320
Db |||||
1959 GCACCGGCTGTGAGGAGGAGGCGCGCAGGAGGAGCAGGCGCACCCCTCTCTGCGCAAGC 2018
QY 1321 CCGCTCATTTGAGACTGTCCCTTCCGGTGTCTGTGCGACCCACTTTGCGCCCTCTGCGCA 1380
Db |||||
2019 CCGCTCATTTGAGACTGTCCCTTCCGGTGTCTGTGCGACCCACTTTGCGCCCTCTGCGCA 2078
QY 1381 CAGCCATCCCTGGAACATGACTCTCGAGCACCCCGCGCCCTCTCTGCGAGGCTGCGG 1440
Db |||||
2079 CAGCCATCCCTGGAACATGACTCTCGAGCACCCCGCGCCCTCTCTGCGAGGCTGCGG 2138
QY 1441 TGAGGCACAGCAGCTGCTTTCAGCCCTTCCGGGGGCGGCCCTTATCAGGAGACATGGGCA 1500
Db |||||
2139 TGAGGCACAGCAGCTGCTTTCAGCCCTTCCGGGGGCGGCCCTTATCAGGAGACATGGGCA 2198
QY 1501 CCTCTAGGGTCTCAAGAGCTTCCATCTGCTGTGCGCACCCAGGCACTGCTCAGCAGA 1560
Db |||||
2199 CCTCTAGGGTCTCAAGAGCTTCCATCTGCTGTGCGCACCCAGGCACTGCTCAGCAGA 2258
QY 1561 GAGGCATCCCGGACAGCCCTTGGGGGAGCAGCCCTTTCTGCGCACCCCAAGCAGG 1620
Db |||||
2259 GAGGCATCCCGGACAGCCCTTGGGGGAGCAGCCCTTTCTGCGCACCCCAAGCAGG 2318
QY 1621 TTCTGCCCCCAGGAGGCTGCGAGCCCCCAGCAGTGTGCCCCCATGCGCTCTCTGGCTC 1680
Db |||||
2319 TTCTGCCCCCAGGAGGCTGCGAGCCCCCAGCAGTGTGCCCCCATGCGCTCTCTGGCTC 2378
QY 1681 CTTTCCCTCAGATCTTGCAAGAGGCCCTTTAGTACCTCAAGCCCTTTCTTGGGACA 1740
Db |||||
2379 CTTTCCCTCAGATCTTGCAAGAGGCCCTTTAGTACCTCAAGCCCTTTCTTGGGACA 2438
QY 1741 GCCCCAGGCACCCCTTGCCTGCCAAGCAAGCCCCCATTTGGACTCTTAAGATGGGGC 1800
Db |||||
2439 GCCCCAGGCACCCCTTGCCTGCCAAGCAAGCCCCCATTTGGACTCTTAAGATGGGGC 2498
QY 1801 TGGAGACATCTCTTCTTGGGAGGCCAAACCCCGGCCCTGCACTTCCCCAGGGTCAAGC 1860
Db |||||
2499 TGGAGACATCTCTTCTTGGGAGGCCAAACCCCGGCCCTGCACTTCCCCAGGGTCAAGC 2558
QY 1861 CTTCCAGGCGAGCTCTTCCCAAGTGTGCTTCTGAGGTGGGCTCTCTCCAGGTGGGAC 1920
Db |||||
2559 CTTCCAGGCGAGCTCTTCCCAAGTGTGCTTCTGAGGTGGGCTCTCTCCAGGTGGGAC 2618
QY 1921 AGAGCTGTGCCCCCTTCCCTGGATGCGAGGCTGGACCCAGAGGCTGAGGATCTCTCCGA 1980
Db |||||
2619 AGAGCTGTGCCCCCTTCCCTGGATGCGAGGCTGGACCCAGAGGCTGAGGATCTCTCCGA 2678
QY 1981 CTTCCACACCCACTTTGCGAGCGGCTCAGGAACAGGCGACCATGGCGCAAGTTCTCCCTGG 2040
Db |||||
2679 CTTCCACACCCACTTTGCGAGCGGCTCAGGAACAGGCGACCATGGCGCAAGTTCTCCCTGG 2738
QY 2041 TGGTCCGGGGGCTACGAGCGCTGGCTGCTATGGCACCTTTGCCCTTTGGTGGAGATGC 2100
Db |||||
2739 TGGTCCGGGGGCTACGAGCGCTGGCTGCTATGGCACCTTTGCCCTTTGGTGGAGATGC 2798
QY 2101 AGGGGCACTGTGGGGCAGGGGCCATGTGGGCCAGGATAGCTGGGCTGTGTCCCACTC 2160
Db |||||
2799 AGGGGCACTGTGGGGCAGGGGCCATGTGGGGCAGGATAGCTGGGCTGTGTCCCACTC 2858
QY 2161 GGAGGAGGAGGAGGAGGAGGCGCGGCTGAGTCCCACTGCGGAGGAGGAGGAGGCG 2220

QY 841 GCACCTCTGCAGGGACACTGTGGCTCTCCAGTTCTCTCTCTCTGACAAAGAGCT 900
DB 1539 GCACCTCTGCAGGGACACTGTGGCTCTCTCCAGTTCTCTCTCTCTCTGACAAAGAGCT 1598
QY 901 CGCCCAATTGCGCGGCTTAAGTCACTGCAACCTCTCCCGGTGACACACTCAACAAGCTGT 960
DB 1599 CGCCCAATTGCGCGGCTTAAGTCACTGCAACCTCTCCCGGTGACACACTCAACAAGCTGT 1658
QY 961 GCACCCCGCGGCTTCTCTCGGCCCTCGGCCAGCTGCTGAGGAAGCGGAGGCCAGTGA 1020
DB 1659 GCACCCCGCGGCTTCTCTCGGCCCTCGGCCAGCTGCTGAGGAAGCGGAGGCCAGTGA 1718
QY 1021 GGGCTCCACGAGGCCCAAGCTCCGCTGCACTCTCCGAGGGTGCAGGCCCAACCGGCCGC 1080
DB 1719 GGGCTCCACGAGGCCCAAGCTCCGCTGCACTCTCCGAGGGTGCAGGCCCAACCGGCCGC 1778
QY 1081 CCAGGGCTGCGTCCCGGCCGACAGCGTCACTCGGACGCTTCTACACACGAGGCGGTGA 1140
DB 1779 CAGGGCTGCGTCCCGGCCGACAGCGTCACTCGGACGCTTCTACACACGAGGCGGTGA 1838
QY 1141 GAGCCCTGAGCACGGGGCCCTTGGCCCGGGGAGCAGCGGCAACCGGCCCGGCCGCA 1200
DB 1839 GAGCCCTGAGCACGGGGCCCTTGGCCCGGGGAGCAGCGGCAACCGGCCCGGCCGCA 1898
QY 1201 CTTGCTGAAGGCGGGTACATTGCGGGGCGGTGCGAGGCTGCGCGAGGCACTGATGGA 1260
DB 1899 CTTGCTGAAGGCGGGTACATTGCGGGGCGGTGCGAGGCTGCGCGAGGCACTGATGGA 1958
QY 1261 GCACCGCTGTGGAGGAGGAGCGCCGAGGAGGAGCAGGCGCAACCGGCCCGGCCGCA 1320
DB 1959 GCACCGCTGTGGAGGAGGAGCGCCGAGGAGGAGCAGGCGCAACCGGCCCGGCCGCA 2018
QY 1321 CCCCTCATTCGAGACTGCCCTCGGCTGCGCTCTGGCAACCGCACTTGGGCCCTGGCCA 1380
DB 2019 CCCCTCATTCGAGACTGCCCTCGGCTGCGCTCTGGCAACCGCACTTGGGCCCTGGCCA 2078
QY 1381 CAGCACTCTCCGGAATGACTCTCGAGACACCGCCCGCCCTCTCGGAGGCTCGGG 1440
DB 2079 CAGCACTCTCCGGAATGACTCTCGAGACACCGCCCGCCCTCTCGGAGGCTCGGG 2138
QY 1441 TGAGGCAAGGCACTGCCCTCAGCGCCCTCGGGGGGCGCCCTATCAGGGACATGCGGCA 1500
DB 2139 TGAGGCAAGGCACTGCCCTCAGCGCCCTCGGGGGGCGCCCTATCAGGGACATGCGGCA 2198
QY 1501 CCCTCAGGGCTCCAAAGCAGCTTCATCCACTGTGTGGCCACCCAGGCACTGCTCAGCCAGA 1560
DB 2199 CCCTCAGGGCTCCAAAGCAGCTTCATCCACTGTGTGGCCACCCAGGCACTGCTCAGCCAGA 2258
QY 1561 GAGGCCATCCCGGACAGCCCTTGGGGGACGCCAGCCCTTCTGCCACCCCAAGCAGGG 1620
DB 2259 GAGGCCATCCCGGACAGCCCTTGGGGGACGCCAGCCCTTCTGCCACCCCAAGCAGGG 2318
QY 1621 TTCTGCCCCCGAGGAGGCTGCAAGCCCGCCACCCAGCAGTTTGGCCCATGCCCTCTGCTC 1680
DB 2319 TTCTGCCCCCGAGGAGGCTGCAAGCCCGCCACCCAGCAGTTTGGCCCATGCCCTCTGCTC 2378
QY 1681 CTTCCTCTCAGGATCTTGCAGAGGCGCCCTTAGTACCTCAAGGCCCTTCTTGGGACA 1740
DB 2379 CTTCCTCTCAGGATCTTGCAGAGGCGCCCTTAGTACCTCAAGGCCCTTCTTGGGACA 2438
QY 1741 GCCCCAGGACCCCTCGCCCTGCAAGCAAGCCCCCATTCGACTCTAAGATGGGGC 1800
DB 2439 GCCCCAGGACCCCTCGCCCTGCAAGCAAGCCCCCATTCGACTCTAAGATGGGGC 2498
QY 1801 TGAGACATCTCTTCTCGGAGGCCAAACCCCGGCCCTGCAAGTTTCCCGAGGTCAGC 1860
DB 2499 TGAGACATCTCTTCTCGGAGGCCAAACCCCGGCCCTGCAAGTTTCCCGAGGTCAGC 2558
QY 1861 CTCCGAGGACCTCTTCCCAAGTACGCTCCCTCAGGTTGGGTCTCTCCAGGTGGGCAC 1920
DB 2559 CTCCGAGGACCTCTTCCCAAGTACGCTCCCTCAGGTTGGGTCTCTCCAGGTGGGCAC 2618
QY 1921 AGAGCTTGGCCCTCTCTCGATGCGAGGGCTGGACCCAGGAGGCTGAGGATCTGTCCGA 1980

DB 2619 AGAGCTTGGCCCTCTCTCGATGCGAGGGCTGACCCAGGAGGCTGAGGATCTGTCCGA 2678
QY 1981 CTCCACACCCACCTTGCAGCGGCTCAGGAACAGGCGACCATGCGAAGTTTCTCCCTGGG 2040
DB 2679 CTCCACACCCACCTTGCAGCGGCTCAGGAACAGGTCACCATGCGAAGTTTCTCCCTGGG 2738
QY 2041 TGGTCGCGGGGCTTACCGAGGCGTGGCTATAGGCACCTTTGCTTGGTGAGATGC 2100
DB 2739 TGGTCGCGGGGCTTACCGAGGCGTGGCTATAGGCACCTTTGCTTGGTGAGATGC 2798
QY 2101 AGGGGGCATGCTGGGGCAGGGGCCCATGTGGGCAGGATAGCCTGTGTCCCAGTC 2160
DB 2799 AGGGGGCATGCTGGGGCAGGGGCCCATGTGGGCAGGATAGCCTGTGTCCCAGTC 2858
QY 2161 GAGGAGGAGGAGCAGGAGGAGGCGCAGGGCTGAGTCCCATGTCGAGGAGCAGCAGGAGGC 2220
DB 2859 GAGGAGGAGGAGCAGGAGGAGGCGCAGGGCTGAGTCCCATGTCGAGGAGCAGCAGGAGGC 2918
QY 2221 CAGGGCTGAGAGCCCACTGCGCCCGAGGTCAGTGCAGGSCCTGTGCTGAGGTCGGCAGGGC 2280
DB 2919 CAGGGCTGAGAGCCCACTGCGCCCGAGGTCAGTGCAGGSCCTGTGCTGAGGTCGGCAGGGC 2978
QY 2281 TCCACACAGGAGCTCTCCAGAGCCCACTGTCGAGGAGCATCGGGCAGGCTCTCCCTGGT 2340
DB 2979 TCCACACAGGAGCTCTCCAGAGCCCACTGTCGAGGAGCATCGGGCAGGCTCTCCCTGGT 3038
QY 2341 GCAGATCCCGGACCTGTGAGGTATGCGGAGGGCGCGCACAAATATCCCTGACATTTTC 2400
DB 3039 GCAGATCCCGGACCTGTGAGGTATGCGGAGGGCGCGCACAAATATCCCTGACATTTTC 3098
QY 2401 CGAGGTGGACCCCGCTTACCTCAACTCTCAGACCTGTACGATATCAAGTACCTCCCATTT 2460
DB 3099 CGAGGTGGACCCCGCTTACCTCAACTCTCAGACCTGTACGATATCAAGTACCTCCCATTT 3158
QY 2461 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCCGTCTCAGCCAGAGCCGCTTCCGCCAT 2520
DB 3159 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCCGTCTCAGCCAGAGCCGCTTCCGCCAT 3218
QY 2521 GGTGAGGAGGAGCTGGCCGAGTTCCGGAGGCCCACTGTCGCCCTGCGCAGGTGAATGGG 2580
DB 3219 GGTGAGGAGGAGCTGGCCGAGTTCCGGAGGCCCACTGTCGCCCTGCGCAGGTGAATGGG 3278
QY 2581 CCCCCACGAGGCTTGGAGATCACAGAGAGTCAGAGGATGTGGACGCGCTGTGTGCAGA 2640
DB 3279 CCCCCACGAGGCTTGGAGATCACAGAGAGTCAGAGGATGTGGACGCGCTGTGTGCAGA 3338
QY 2641 GGTGCGGTGGGAGGAGGCGCAAGTGTCTCGCCGTCAAGCAGCCTTTCACACTTCCC 2700
DB 3339 GGTGCGGTGGGAGGAGGCGCAAGTGTCTCGCCGTCAAGCAGCCTTTCACACTTCCC 3398
QY 2701 TGGGAGGACCTTGGCGGTGAGCTGAGGCTGGGGCTGCGTGAGAGGTGAGGC 2760
DB 3399 TGGGAGGACCTTGGCGGTGAGCTGAGGCTGGGGCTGCGTGAGAGGTGAGGC 3458
QY 2761 CTCCGTGGAGCACATCTCCCGGATCCTGAAGGGCAGGCCGGAAGGTCTCGAGAGGAGG 2820
DB 3459 CTCCGTGGAGCACATCTCCCGGATCCTGAAGGGCAGGCCGGAAGGTCTCGAGAGGAGG 3518
QY 2821 GCGCCCCAGGAAGACAGGCTTGTCTTTCCTTCCGGCTCTCAGGTCTGAAAGGTGGGA 2880
DB 3519 GCGCCCCAGGAAGACAGGCTTGTCTTTCCTTCCGGCTCTCAGGTCTGAAAGGTGGGA 3578
QY 2881 CCGAGCGCCGACATTTCTAAGGGAGCTCTCAGATGAGACTGTGGTCTTGGGCCAGTCACT 2940
DB 3579 CCGAGCGCCGACATTTCTAAGGGAGCTCTCAGATGAGACTGTGGTCTTGGGCCAGTCACT 3638
QY 2941 GACATGCGCTGCCAGGTGTGAGCCAGCAGCTGCCAGGCACTCGAGCAAGACGG 3000
DB 3639 GACATGCGCTGCCAGGTGTGAGCCAGCAGCTGCCAGGCACTCGAGCAAGACGG 3698
QY 3001 A 3001

Db 3699 A 3699

RESULT 13

US-10-921-168-1

; Sequence 1, Application US/10921168

; Publication No. US2005000346A1

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui, et al

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL000927-CIP-DIV3

; CURRENT APPLICATION NUMBER: US/10/921.168

; CURRENT FILING DATE: 2004-08-19

; PRIOR FILING DATE: 10/274,978

; PRIOR FILING DATE: 2002-10-22

; PRIOR APPLICATION NUMBER: 09/858,664

; PRIOR FILING DATE: 2001-05-17

; PRIOR APPLICATION NUMBER: 09/711,134

; PRIOR FILING DATE: 2000-11-14

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 5207

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-921-168-1

Query Match 99.9%; Score 2997.8; DB 18; Length 5207;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCTACGGAGAGAACTCGGGCCAGGACATACAGGAGGAGAGACATCTCTGGCGGGCTGAG 60

Db 699 CCTACGGAGAGAACTCGGGCCAGGACATACAGGAGGAGAGACATCTCTGGCGGGCTGAG 758

Qy 61 CCACCGGTGTACGGGGCTGTGGACAGTTTGGACCCGGAAGACCTCATCTCTCAT 120

Db 759 CCACCGGTGTACGGGGCTGTGGACAGTTTGGACCCGGAAGACCTCATCTCTCAT 818

Qy 121 CCTGGAGCTGTCTATCCGAGGAGCTGTGGACCGCTGTACAGGAGGGCGTGGTAC 180

Db 819 CCTGGAGCTGTCTATCCGAGGAGCTGTGGACCGCTGTACAGGAGGGCGTGGTAC 878

Qy 181 GGAGGCCAGGTCAAGGTCTATCATCCAGCAGCTGTGGAGGGGCTGCACTACCTGCAAG 240

Db 879 GGAGGCCAGGTCAAGGTCTATCATCCAGCAGCTGTGGAGGGGCTGCACTACCTGCAAG 938

Qy 241 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTAAACATCTGTATGGTGCATCTTGGCCG 300

Db 939 CCATGGCGTTCTCCACCTGGACATAAAGCCCTCTAAACATCTGTATGGTGCATCTTGGCCG 998

Qy 301 GGAAGACATTAATCTGCGACTTTGGCTTTGGCCAGAACATCAACCCAGCAGAGCTGCA 360

Db 999 GGAAGACATTAATCTGCGACTTTGGCTTTGGCCAGAACATCAACCCAGCAGAGCTGCA 1058

Qy 361 GTTCAGCCAGTACCGCTCCCTGAGTTGCTCTCCCGGAGATCATCCAGCAGAACCTGT 420

Db 1059 GTTCAGCCAGTACCGCTCCCTGAGTTGCTCTCCCGGAGATCATCCAGCAGAACCTGT 1118

Qy 421 GAGCGAAGCCTCCGACATTTGGGCGATGGGTGTCTATCTCTCACTCTGAGCCTGACTGCTC 480

Db 1119 GAGCGAAGCCTCCGACATTTGGGCGATGGGTGTCTATCTCTCACTCTGAGCCTGACTGCTC 1178

Qy 481 ATCCCCATTTCCCGGAGAGTACCGGTGGCCACCTCTGAACTCTCTGGAGGGCGCGT 540

Db 1179 ATCCCCATTTCCCGGAGAGTACCGGTGGCCACCTCTGAACTCTCTGGAGGGCGCGT 1238

Qy 541 GTCATGAGCAGCCCCCATGGCTGCCCCACTCAGCGAAGACGCCAAGACTTTTCATCAAGGC 600

Db 1239 GTCATGAGCAGCCCCCATGGCTGCCCCACTCAGCGAAGACGCCAAGACTTTTCATCAAGGC 1298

Qy 601 TAGCTGCAGAGAGCCCTCAGGCCCGGCTAGTGGGCGCCAGTGCCTCTCTCCACCCCTG 660

Db 1299 TAGCTGCAGAGAGCCCTCAGGCCCGGCTAGTGGGCGCCAGTGCCTCTCTCCACCCCTG 1358

Qy 661 GTTCTGAAATCCATGCTCGGAGGAGGGCCACTTTCATCAACACCAAGCAGCTCAAGTT 720

Db 1359 GTTCTGAAATCCATGCTCGGAGGAGGGCCACTTTCATCAACACCAAGCAGCTCAAGTT 1418

Qy 721 CTTCTGGCCGAAAGTCTGTGGGAGGCTTCTCTGATGAGCTTACAGTTCATCTCTGGTGTAT 780

Db 1419 CTTCTGGCCGAAAGTCTGTGGGAGGCTTCTCTGATGAGCTTACAGTTCATCTCTGGTGTAT 1478

Qy 781 GCGCTCCATCCCTGAGCTGTGGGGGCCACACCGACAGCCCTCTCTCGGCTAGCCCG 840

Db 1479 GCGCTCCATCCCTGAGCTGTGGGGGCCACACCGACAGCCCTCTCTCGGCTAGCCCG 1538

Qy 841 GCACCTCTGCAGGAGACACTGTGTGCTCTCTCCAGTTCTCTCTCTCTGACAAAGAGCT 900

Db 1539 GCACCTCTGCAGGAGACACTGTGTGCTCTCTCCAGTTCTCTCTCTCTGACAAAGAGCT 1598

Qy 901 GCGCCCATTTGCGGGCTAAGTCACTGCGACCCCTCCCGGTGACACTCACCAGCT 960

Db 1599 GCGCCCATTTGCGGGCTAAGTCACTGCGACCCCTCCCGGTGACACTCACCAGCT 1658

Qy 961 GCACCCCGGGGCTTCTGTGGGCCCTCGGCCAGCTGCTGAGGAGCCGAGGCGCAGTGA 1020

Db 1659 GCACCCCGGGGCTTCTGTGGGCCCTCGGCCAGCTGCTGAGGAGCCGAGGCGCAGTGA 1718

Qy 1021 GCGCTCCACGAGGCCCGCAGCTCCGCTGTCTCTCCGAGGGTCCGCGGCCACCGGCGCG 1080

Db 1719 GCGCTCCACGAGGCCCGCAGCTCCGCTGTCTCTCCGAGGGTCCGCGGCCACCGGCGCG 1778

Qy 1081 CCAGGGTGTGTGCGCGGACAGCGTCACTCCGAGCTGTTCACACAGGCGGGTGA 1140

Db 1779 CCAGGGTGTGTGCGCGGACAGCGTCACTCCGAGCTGTTCACACAGGCGGGTGA 1838

Qy 1141 GAGCCCTGAGCAGGGGCCCTTGGCCCGGGGAGCAGGGGACCCCGCCCGCGCGCGCA 1200

Db 1839 GAGCCCTGAGCAGGGGCCCTTGGCCCGGGGAGCAGGGGACCCCGCGCGCGCGCGCA 1898

Qy 1201 CCTCTGAAGGGCGGTACATTGGGGGGGCTCTCCAGGCTCGCGGACCACTGATGGA 1260

Db 1899 CCTCTGAAGGGCGGTACATTGGGGGGGCTCTCCAGGCTCGCGGACCACTGATGGA 1958

Qy 1261 GCACCGGTGTGAGAGAGAGCGCGCAGGAGGAGCAGGCCACCTCTCTGGCCAAAGC 1320

Db 1959 GCACCGGTGTGAGAGAGAGCGCGCAGGAGGAGCAGGCCACCTCTCTGGCCAAAGC 2018

Qy 1321 CCCCTCATTCGAGACTGCCCTCGGCTGCTCTGGCACCCACTTGGGCCCTTGGCGCA 1380

Db 2019 CCCCTCATTCGAGACTGCCCTCGGCTGCTCTGGCACCCACTTGGGCCCTTGGCGCA 2078

Qy 1381 CAGCCACTCTCTGAAACATGACTCTCCGAGCAGCCCCCGCCCTCTCTCGAGGGCTGCGG 1440

Db 2079 CAGCCACTCTCTGAAACATGACTCTCCGAGCAGCCCCCGCCCTCTCTCGAGGGCTGCGG 2138

Qy 1441 TGAGGCA CAGCGACTGCTTTCAGCCCTCTCGGGGGGCGCCCTATTCAGGGACATGGGGCA 1500

Db 2139 TGAGGCA CAGCGACTGCTTTCAGCCCTCTCGGGGGGCGCCCTATTCAGGGACATGGGGCA 2198

Qy 1501 CCCTCAGGGTCCAGCAGCTTCCATCCATGGTGGCCACCCAGGCACTGTCTAGCCAGA 1560

Db 2199 CCCTCAGGGTCCAGCAGCTTCCATCCATGGTGGCCACCCAGGCACTGTCTAGCCAGA 2258

Qy 1561 GAGGCCATCCCGGAGCAGCCCTTGGGGGAGCAGCAGCCCTTCTTCTGCCACCCCAAGCAGG 1620

Db 2259 GAGGCCATCCCGGAGCAGCCCTTGGGGGAGCAGCAGCCCTTCTTCTGCCACCCCAAGCAGG 2318

Qy 1621 TTCTGCCCGCCAGGAGGGCTGACGCCCCCAACCCAGCAGTTTGGCCCATCTCTCTGGCTC 1680

Db 2319 TTCTGCCCGCCAGGAGGGCTGACGCCCCCAACCCAGCAGTTTGGCCCATCTCTCTGGCTC 2378

Qy 1681 CTTCCCTCAGGATCTTGCAGAGGGGCCCTTAGTACCTCTAAGCCCTTCTTGGGACA 1740

Db 2379 CTTCCCTCCAGGATCTTGAAGAGGCCCTTAGTACCTCAAGCCCCCTTCTGGGACA 2438
QY 1741 GCCCCAGGACACCCCTGCCCCCTGCCAAAGCAAGCCCCCATTTGGAATCTTAAGATGGGCC 1800
Db 2439 GCCCCAGGACACCCCTGCCCCCTGCCAAAGCAAGCCCCCATTTGGAATCTTAAGATGGGCC 2498
QY 1801 TGGAGACATCTCTCTTCTTCTGGAGGCCAAAAACCCCGCCCCCTGCAGTTCCCCAGGCTCAGC 1860
Db 2499 TGGAGACATCTCTCTTCTTCTGGAGGCCAAAAACCCCGCCCCCTGCAGTTCCCCAGGCTCAGC 2558
QY 1861 CTCCTCAGGCGAGCTCTTCCCAAGTGAGCTCCCTCAGGCTGGGCTCTCTCCAGGTGGGAC 1920
Db 2559 CTCCTCAGGCGAGCTCTTCCCAAGTGAGCTCCCTCAGGCTGGGCTCTCTCCAGGTGGGAC 2618
QY 1921 AGAGCCTGGCCCCCTCCTCGATGCGGAGGCTGGACCCAGGAGGCTGAGGATCTCTCCGA 1980
Db 2619 AGAGCCTGGCCCCCTCCTCGATGCGGAGGCTGGACCCAGGAGGCTGAGGATCTCTCCGA 2678
QY 1981 CTCACACCCACTTGTGAGCGGCTCGAGAAACAGCGACCATGCGCAAGTTCTCCCTGGG 2040
Db 2679 CTCACACCCACTTGTGAGCGGCTCGAGAAACAGGTCACCATGCGCAAGTTCTCCCTGGG 2738
QY 2041 TGGTCCGCGGGCTACGACAGGCTGGCTGGCTATGCGACCTTTGCTTTGTTGAGATGC 2100
Db 2739 TGGTCCGCGGGCTACGACAGGCTGGCTGGCTATGCGACCTTTGCTTTGTTGAGATGC 2798
QY 2101 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGTCCAGTC 2160
Db 2799 AGGGGGCATGTGGGGCAGGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGTCCAGTC 2858
QY 2161 GGAGAGAGAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2220
Db 2859 GGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2918
QY 2221 CAGGCTCAGAGCCACTGCCACAGTCAAGTCAAGGCTGTGCTCAGGTCGAGGAGGAGG 2280
Db 2919 CAGGCTCAGAGCCACTGCCACAGTCAAGTCAAGGCTGTGCTCAGGTCGAGGAGGAGG 2978
QY 2281 TCCACACAGGAGCTCTCCAGAGCCACCCCATGGAGGACATCGGGCAGGCTCTCCCTGGT 2340
Db 2979 TCCACACAGGAGCTCTCCAGAGCCACCCCATGGAGGACATCGGGCAGGCTCTCCCTGGT 3038
QY 2341 CGAGATCGGACCTGTCAAGTGTATGCGGAGGCGGCGGCGGACACAATATCCCTGGACATTC 2400
Db 3039 CGAGATCGGACCTGTCAAGTGTATGCGGAGGCGGCGGCGGACACAATATCCCTGGACATTC 3098
QY 2401 CGAGTGGACCCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db 3099 CGAGTGGACCCCGCTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 3158
QY 2461 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCGAGAGCCCGCTCCCCAT 2520
Db 3159 CGAGTTTATGATCTTCAGGAAAGTCCCAAGTCCGCTCAGCGAGAGCCCGCTCCCCAT 3218
QY 2521 GGCTGAGAGAGCTGGCGAGTTCCCGAGGCCCGGAGGCGGCTGGCGGCTGGCGAGTGAATGGG 2580
Db 3219 GGCTGAGAGAGCTGGCGAGTTCCCGAGGCCCGGAGGCGGCTGGCGGCTGGCGAGTGAATGGG 3278
QY 2581 CCCCCACGAGGCTGGAGATCAGAGAGGAGTCAAGAGATGTGACGCGCTGTGGCAGA 2640
Db 3279 CCCCCACGAGGCTGGAGATCAGAGAGGAGTCAAGAGATGTGACGCGCTGTGGCAGA 3338
QY 2641 GGCTGCGGTGGGAGGAGGAGGAGTGTGCTCGCGCTCAGCGGCTCTTCCACTTCCC 2700
Db 3339 GGCTGCGGTGGGAGGAGGAGGAGTGTGCTCGCGCTCAGCGGCTCTTCCACTTCCC 3398
QY 2701 TGGAGGACACTGCGCGTGTAGCTGAGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 2760
Db 3399 TGGAGGACACTGCGCGTGTAGCTGAGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 3458
QY 2761 CTCGTGGAGACATCTCCCGATCTTGAAAGGCGAGGCCGGAAGGCTCTGGAGAGGAGG 2820

Db 3459 CTCCTGGAGCACATCTCCCGATCTCAAGGCGAGGCCGGAAGGTCTGGAGAAGGAGG 3518
QY 2821 GCCCCCCAGGAGAGGAGCCAGGCTTGTCTTCCGGCTCTCAGGTCTGAAGAGCTGGGA 2880
Db 3519 GCCCCCCAGGAGAGGAGCCAGGCTTGTCTTCCGGCTCTCAGGTCTGAAGAGCTGGGA 3578
QY 2881 CCGAGCGCCGACATCTCTAAGGGAGCTCTCAGATGAGACTGTGTCTCTGGGCCAGTCA 2940
Db 3579 CCGAGCGCCGACATCTCTAAGGGAGCTCTCAGATGAGACTGTGTCTCTGGGCCAGTCA 3638
QY 2941 GACACTGGCTGCGCAGGTGTGAGCCAGGCTGAGCCAGGCTGAGCCAGGCTGAGGAGGAGG 3000
Db 3639 GACACTGGCTGCGCAGGTGTGAGCCAGGCTGAGCCAGGCTGAGGAGGAGGAGGAGG 3698
QY 3001 A 3001
Db 3699 A 3699

RESULT 14
US-10-921-168-3
; Sequence 3, Application US/10921168
; Publication No. US20050003446A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV3
; CURRENT APPLICATION NUMBER: US/10/921,168
; PRIOR FILING DATE: 2004-08-19
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-921-168-3

Query Match 99.9%; Score 2997.8; DB 18; Length 5207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTACGGAGCAGAACTCGGGCCAGGCATACAGGAGGAGGAGACATCTCTGGCCGCGTGAG 60
Db 699 CTTACGGAGCAGAACTCGGGCCAGGCATACAGGAGGAGGAGACATCTCTGGCCGCGTGAG 758
QY 61 CCACCCGCTGGTCAAGGCTCTGAGCCAGTTTGGAGCCCGCAAGACCTCATCTCAT 120
Db 759 CCACCCGCTGGTCAAGGCTCTGAGCCAGTTTGGAGCCCGCAAGACCTCATCTCAT 818
QY 121 CTTGAGCTGTCTCATCCGAGGAGCTGTGAGCCGCTGTACAGGAGGAGGAGGAGGAGGAG 180
Db 819 CTTGAGCTGTCTCATCCGAGGAGCTGTGAGCCGCTGTACAGGAGGAGGAGGAGGAGGAG 878
QY 181 GGAGCCGAGGTCAGAGTCTACATCCAGAGCTGTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
Db 879 GGAGCCGAGGTCAGAGTCTACATCCAGAGCTGTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 938
QY 241 CCATGGCTTCTCCACCTGGACATTAAGCCCTCTAAACATCTGATGGTGCATCTCTGCCCG 300
Db 939 CCATGGCTTCTCCACCTGGACATTAAGCCCTCTAAACATCTGATGGTGCATCTCTGCCCG 998
QY 301 GGAAGACATTAATAATCTCGGACTTTGGCTTTGGCTTTGGCCAGAACATCACCCAGCAGAGCTGCA 360
Db 999 GGAAGACATTAATAATCTCGGACTTTGGCTTTGGCCAGAACATCACCCAGCAGAGCTGCA 1058

QY 361 GTTCAGCAGTACGGCTCCCTGAGTTGCTCTCCCGAGATCATCCAGCAGAACCCCTGT 420
Db 1059 GTTCAGCAGTACGGCTCCCTGAGTTGCTCTCCCGAGATCATCCAGCAGAACCCCTGT 1118
QY 421 GAGCGAAGCCTCCGACATATTTGGGCCATGGGTGTCTCTTACCTCAGACCTGACCTGCTC 480
Db 1119 GAGCGAAGCCTCCGACATATTTGGGCCATGGGTGTCTCTTACCTCAGACCTGACCTGCTC 1178
QY 481 ATCCCCATTTGCCGGCAGAGTACGGTGCCTGACACCTCTCTGAACTCTCTGAGGGGGCGGT 540
Db 1179 ATCCCCATTTGCCGGCAGAGTACGGTGCCTGACACCTCTCTGAACTCTCTGAGGGGGCGGT 1238
QY 541 GTCATGAGCAGCAGCCCATGGCTGCCACCTCAGCGAAGACGCCAAAGACTTTCATCAAGGC 600
Db 1239 GTCATGAGCAGCAGCCCATGGCTGCCACCTCAGCGAAGACGCCAAAGACTTTCATCAAGGC 1298
QY 601 TAGCCTCAGAGAGCCCTCAGGCCCGGCTAGTGGCGGCCAGTGCCTCTCCCAACCCCTG 660
Db 1299 TAGCCTCAGAGAGCCCTCAGGCCCGGCTAGTGGCGGCCAGTGCCTCTCCCAACCCCTG 1358
QY 661 GTTCCTGAATCCATGCTTGGGAGGAGGCCACCTTCATCAACCAAGCAGCTCAAGTT 720
Db 1359 GTTCCTGAATCCATGCTTGGGAGGAGGCCACCTTCATCAACCAAGCAGCTCAAGTT 1418
QY 721 CCTCTGGCCCGAAGTCTGGCAGCGTTCCCTCATGAGCTACAAGTCCATCTCTGGTGAT 780
Db 1419 CCTCTGGCCCGAAGTCTGGCAGCGTTCCCTCATGAGCTACAAGTCCATCTCTGGTGAT 1478
QY 781 GCGCTCATCTCCTGAGCTGTGCGGGGCCCAACCGACAGCCCTCCCTCGGCGTAGCCCG 840
Db 1479 GCGCTCATCTCCTGAGCTGTGCGGGGCCCAACCGACAGCCCTCCCTCGGCGTAGCCCG 1538
QY 841 GCACCTTCAGAGGACACTGGTGGCTCTCAGTTCTCTCTCTCTCTCTCTCTGACCAAGAGCT 900
Db 1539 GCACCTTCAGAGGACACTGGTGGCTCTCAGTTCTCTCTCTCTCTCTCTCTGACCAAGAGCT 1598
QY 901 CGCCCCATTTGCCCGGCTAAGTCACTGCCACCTCCCGGTGACACTCACCACCTGCT 960
Db 1599 CGCCCCATTTGCCCGGCTAAGTCACTGCCACCTCCCGGTGACACTCACCACCTGCT 1658
QY 961 GCACCCCCGGGGTTCCTCGGGCCCTCGGGCAGCTGCTGAGGAAGCCGAGGCGCAGTGA 1020
Db 1659 GCACCCCCGGGGTTCCTCGGGCCCTCGGGCAGCTGCTGAGGAAGCCGAGGCGCAGTGA 1718
QY 1021 GCGCTTCACGAGGCCCTCGGCTCGCTGCTATCTCCGAGGGTCCGGGCCACCGGGCGC 1080
Db 1719 GCGCTTCACGAGGCCCTCGGCTCGCTGCTATCTCCGAGGGTCCGGGCCACCGGGCGC 1778
QY 1081 CCAGGGTGTGTCGCCCGGCACAGCTCATCCGAGAGCTGTTCTACCAAGGCGGGTGA 1140
Db 1779 CCAGGGTGTGTCGCCCGGCACAGCTCATCCGAGAGCTGTTCTACCAAGGCGGGTGA 1838
QY 1141 GAGCCTCAGCAGGGGCCCTTGGCCCGGGAGCAGGGCGCACCGCGCCCGGGCGGCGCA 1200
Db 1839 GAGCCTCAGCAGGGGCCCTTGGCCCGGGAGCAGGGCGCACCGCGCCCGGGCGGCGCA 1898
QY 1201 CTGCTGAAGGGCGGGTACATTGGGGGGCGCTGCCAGGCCCTGCGGAGCCACTGTATGGA 1260
Db 1899 CTGCTGAAGGGCGGGTACATTGGGGGGCGCTGCCAGGCCCTGCGGAGCCACTGTATGGA 1958
QY 1261 GCACCGGTGTGAGGAGGAGGGCGCCAGGGAGGAGCGGCCACCTCTCTGGGCCAAGC 1320
Db 1959 GCACCGGTGTGAGGAGGAGGGCGCCAGGGAGGAGCGGCCACCTCTCTGGGCCAAGC 2018
QY 1321 CCCCTCATTTGAGACTGCCCTCGGGCTGCTCTGAGCAGCCACCTTTGGGCCCTGGCGCA 1380
Db 2019 CCCCTCATTTGAGACTGCCCTCGGGCTGCTCTGAGCAGCCACCTTTGGGCCCTGGCGCA 2078
QY 1381 CAGCCATCTCCTGAAATGATCTCCGAGCAACCCCGGCCCTCTCTCGAGGGCCTCGG 1440
Db 2079 CAGCCATCTCCTGAAATGATCTCCGAGCAACCCCGGCCCTCTCTCGAGGGCCTCGG 2138
QY 1441 TGAGGCACAGGACTGCTTCAGCCCCCTCGGGGGGGCCCCCTATCAGGAGNATCGGGCA 1500

Db 2139 TGAGGCACAGGACTGCTTCAGCCCCCTCGGGGGGGCCCCCTATCAGGAGNATCGGGCA 2198
QY 1501 CCCTCAGGGTCCAAAGCAGCTTCCATCCACTGTGTGGCCACCCAGGCACTGTCTCAGGCCAGA 1560
Db 2199 CCCTCAGGGTCCAAAGCAGCTTCCATCCACTGTGTGGCCACCCAGGCACTGTCTCAGGCCAGA 2258
QY 1561 GAGGCCATCCCGGACAGCCCTTGGGGGCGAGCCAGCCCTTCTGCCACCCCAAGCAGGG 1620
Db 2259 GAGGCCATCCCGGACAGCCCTTGGGGGCGAGCCAGCCCTTCTGCCACCCCAAGCAGGG 2318
QY 1621 TTCTGCCCCCAGGAGGGCTGACGCCCCCAAGCAGAGTTGCCCCATGCCCTCTCTGGCTC 1680
Db 2319 TTCTGCCCCCAGGAGGGCTGACGCCCCCAAGCAGAGTTGCCCCATGCCCTCTCTGGCTC 2378
QY 1681 CTTCCTCTCAGGATCTTTGCAAGAGGCCCTTAGTACCTCAAGCCCTTCTTTGGGACA 1740
Db 2379 CTTCCTCTCAGGATCTTTGCAAGAGGCCCTTAGTACCTCAAGCCCTTCTTTGGGACA 2438
QY 1741 GCCCCAGGCAACCCCTGCGCTGCCAAGAGGCCCCCTTGGACTCTAAGATGGGGCC 1800
Db 2439 GCCCCAGGCAACCCCTGCGCTGCCAAGAGGCCCCCTTGGACTCTAAGATGGGGCC 2498
QY 1801 TGAGAGACTCTCTCTTCTGGGAGGCCAAACCCGGGCCCTGCAAGTTCCCCCAGGGTCAGC 1860
Db 2499 TGAGAGACTCTCTCTTCTGGGAGGCCAAACCCGGGCCCTGCAAGTTCCCCCAGGGTCAGC 2558
QY 1861 CTCCAGGCGAGCTTTCCAAAGTAGCTCTCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 1920
Db 2559 CTCCAGGCGAGCTTTCCAAAGTAGCTCTCTCAGGGTGGGCTCTCTCCAGGTGGGCAC 2618
QY 1921 AGAGCCTGGCCCCCTCTGATGCGGAGGGCTGAGCCAGGAGCTGAGGATCTGTCCGA 1980
Db 2619 AGAGCCTGGCCCCCTCTGATGCGGAGGGCTGAGCCAGGAGCTGAGGATCTGTCCGA 2678
QY 1981 CTCCACACCACTTTCAGCGGCTCAGGAAACAGGCGACCATGCGCAAGTTCTCCCTGGG 2040
Db 2679 CTCCACACCACTTTCAGCGGCTCAGGAAACAGGCGACCATGCGCAAGTTCTCCCTGGG 2738
QY 2041 TGCTCGGGGGGTACCCAGCGGCTGTGGCTATGCGACCTTTGCTTGTGTGAGATGC 2100
Db 2739 TGCTCGGGGGGTACCCAGCGGCTGTGGCTATGCGACCTTTGCTTGTGTGAGATGC 2798
QY 2101 AGGGGGCATCTGGGGCAGGGGCCCATGCGGCAGGATAGCCTGGGCTGTGTCCAGTC 2160
Db 2799 AGGGGGCATCTGGGGCAGGGGCCCATGCGGCAGGATAGCCTGGGCTGTGTCCAGTC 2858
QY 2161 GGAGGAGGAGGAGGAGGAGGAGGCGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 2220
Db 2859 GGAGGAGGAGGAGGAGGAGGAGGCGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 2918
QY 2221 CAGGGCTGAGAGGCCACTGCCCCCAGGTCAAGTCAAGCCCTGTGTGAGTTCGAGGCGAGGC 2280
Db 2919 CAGGGCTGAGAGGCCACTGCCCCCAGGTCAAGTCAAGCCCTGTGTGAGTTCGAGGCGAGGC 2978
QY 2281 TCCACACAGGAGCTCTCCAGAGCCCAACCCATGCGGAGGACATCGGGCAGGTCTCCCTGGT 2340
Db 2979 TCCACACAGGAGCTCTCCAGAGCCCAACCCATGCGGAGGACATCGGGCAGGTCTCCCTGGT 3038
QY 2341 GCAGATCCGGGACCTGTGAGTGTGAGGGGGCGGCGACACATATCCCTGGGACATTTTC 2400
Db 3039 GCAGATCCGGGACCTGTGAGTGTGAGGGGGCGGCGACACATATCCCTGGGACATTTTC 3098
QY 2401 CGAGGTGGACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db 3099 CGAGGTGGACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 3158
QY 2461 CGAGTTTATGATTTTCAAGGAAAGTCCCCAAGTCCGCTCAGCAGAGCGGCCCTCCCCAT 2520
Db 3159 CGAGTTTATGATTTTCAAGGAAAGTCCCCAAGTCCGCTCAGCAGAGCGGCCCTCCCCAT 3218
QY 2521 GGCTGAGGAGGAGCTGGCGGAGTTCCCGGAGCCACCGTGGCCCTGGCCAGGTGAATGGG 2580

Db 3219 GGTGAGGAGGAGCTGGCCGAGTTCCCGAGGCCACGTGGCCCTGGCCAGAGTGAACCTGGG 3278
QY 2581 CCCCCACGAGGCTGGAGATCACAGAGGAGTGGACGGCTGCGCGAGA 2640
Db 3279 CCCCCACGAGGCTGGAGATCACAGAGGAGTGGACGGCTGCGCGAGA 3338
QY 2641 GGTGCGGTGGGAGGAGGCGAAGTGGTCTCGCGGTCAAGAGCTTCCACTTCCC 2700
Db 3339 GGTGCGGTGGGAGGAGGCGAAGTGGTCTCGCGGTCAAGAGCTTCCACTTCCC 3398
QY 2701 TGGAGGACCTGCGCTGGATGAGCTGAGAGTGGGGCTGCGTGGAGAGTGAAGGC 2760
Db 3399 TGGAGGACCTGCGCTGGATGAGCTGAGAGTGGGGCTGCGTGGAGAGTGAAGGC 3458
QY 2761 CTCCTGGAGACATCTCCGGATCTGAAGGCGAGGCGGAAAGTCTGGAGAGGAGGG 2820
Db 3459 CTCCTGGAGACATCTCCGGATCTGAAGGCGAGGCGGAAAGTCTGGAGAGGAGGG 3518
QY 2821 GCCCCCCAGAGAGCCAGGCTTGGTCTTCTCCGGTCTCAGTCTGAAGAGTGGGA 2880
Db 3519 GCCCCCCAGAGAGCCAGGCTTGGTCTTCTCCGGTCTCAGTCTGAAGAGTGGGA 3578
QY 2881 CCGAGCGCCGACATCTTAAGGAGCTCTCAGATGAGAGTGGTCTCGGCTCAGTCTGAGAGTGGGA 2940
Db 3579 CCGAGCGCCGACATCTTAAGGAGCTCTCAGATGAGAGTGGTCTCGGCTCAGTCTGAGAGTGGGA 3638
QY 2941 GACACTGGCTCCAGGTGTGAGCCAGCCAGCTGCGGAGCCAGCTGGAGCAAGAGCGG 3000
Db 3639 GACACTGGCTCCAGGTGTGAGCCAGCCAGCTGCGGAGCCAGCTGGAGCAAGAGCGG 3698
QY 3001 A 3001
Db 3699 A 3699

RESULT 15

US-10-415-011-44
; Sequence 44, Application US/10415011
; Publication No. US20040053394A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: DING, Li
; APPLICANT: TANG, Y. Tom
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: NGUYEN, Dannie B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LU, Yan
; APPLICANT: YUE, Henry
; APPLICANT: BURFORD, Neil
; APPLICANT: BANDMAN, Olga
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: LAL, Preeti G.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: SHARMAKER Anita
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: KHAN, Farrah A.
; APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0262 USN
; CURRENT APPLICATION NUMBER: US/10/415,011
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/47728

Query Match 99.8%; Score 2996.2; DB 17; Length 5454;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCTACGGAGCAGAACTCGGGCCCGAGGATACAGGAGGAGGAGATCCTTGGCGCGCTGAG 60
Db 963 CCTACGGAGCAGAACTCGGGCCCGAGGATACAGGAGGAGGAGATCCTTGGCGCGCTGAG 1022
QY 61 CACCCCGCTGGTCA CGGGGCTGCTGAG CCAAGTTGAGACCCCAAGACCTTCATCTCAT 120
Db 1023 CCACCCGCTGGTCA CGGGGCTGCTGAG CCAAGTTGAGACCCCAAGACCTTCATCTCAT 1082
QY 121 CTTGAGAGCTGTCTCATCCGAGGAGCTGTGACCGCTGTACAGAGGAGGAGGCTGGTGAC 180
Db 1083 CTTGAGAGCTGTCTCATCCGAGGAGCTGTGACCGCTGTACAGAGGAGGAGGCTGGTGAC 1142
QY 181 GGAGGCCGAGGTCAAGGTCTACATCCAGCAGCTGGTGGAGGGGCTGCATCTCTGCACAG 240
Db 1143 GGAGGCCGAGGTCAAGGTCTACATCCAGCAGCTGGTGGAGGGGCTGCATCTCTGCACAG 1202
QY 241 CCATGGCTTCTCCACCTGGACATTAAGCCCTCTAACATCTGTATGGTGGTGCATCTTGCCTG 300
Db 1203 CCATGGCTTCTCCACCTGGACATTAAGCCCTCTAACATCTGTATGGTGGTGCATCTTGCCTG 1262
QY 301 GGAAGACATTTAAATCTCGACTTTGGCTTTGGCCCGAGAACATCACCCAGAGAGCTGCA 360
Db 1263 GGAAGACATTTAAATCTCGACTTTGGCTTTGGCCCGAGAACATCACCCAGAGAGCTGCA 1322
QY 361 GTTCAGCAGTACGGCTCCCTGAGTTCTCTCCCGAGATCATCCAGAGAACCTCTGT 420
Db 1323 GTTCAGCAGTACGGCTCCCTGAGTTCTCTCCCGAGATCATCCAGAGAACCTCTGT 1382
QY 421 GAGCGAAGCCTCCGACATTTGGGCCATGGGTGTCTCTCTACCTCAGCTGACCTGCTC 480
Db 1383 GAGCGAAGCCTCCGACATTTGGGCCATGGGTGTCTCTCTACCTCAGCTGACCTGCTC 1442
QY 481 ATCCCATTTTCCGCGAGAGTGCAGCGTCCACCTCTCTGAACCTCTCTGGAGGGGCGGT 540
Db 1443 ATCCCATTTTCCGCGAGAGTGCAGCGTCCACCTCTCTGAACCTCTCTGGAGGGGCGGT 1502
QY 541 GTCATGGAGAGCCCATGGCTGCCACCTCAGCGAGAGCCCAAGACTTCATCAAGGC 600
Db 1503 GTCATGGAGAGCCCATGGCTGCCACCTCAGCGAGAGCCCAAGACTTCATCAAGGC 1562
QY 601 TACGCTGAGAGAGCCCTCAGGCGCCGCGCTAGTGGCGCCGAGTGCCTCTCCACCCCTG 660
Db 1563 TACGCTGAGAGAGCCCTCAGGCGCCGCGCTAGTGGCGCCGAGTGCCTCTCCACCCCTG 1622

QY 661 GTTCCTGAATCCATGCTGCGAGGAGGCCACTTATCAACCAACAGAGAGCTCAAGTT 720
Db 1623 GTTCCTGAATCCATGCTGCGAGGAGGCCACTTATCAACCAACAGAGAGCTCAAGTT 1682
QY 721 CCTCTGCGCCGAAAGTGGCTGGGACAGGTTCCCTGATGAGCTACAAGTCCATCTCTGGTGAT 780
Db 1683 CCTCTGCGCCGAAAGTGGCTGGGACAGGTTCCCTGATGAGCTACAAGTCCATCTCTGGTGAT 1742
QY 781 GCGTTCATCCCTGAGCTGTGCGGGGCCCAACCGACAGCCCTCCCTCGCGGTAGGCCG 840
Db 1743 GCGTTCATCCCTGAGCTGTGCGGGGCCCAACCGACAGCCCTCCCTCGCGGTAGGCCG 1802
QY 841 GCACCTCTGACGAGGACACTGAGTCACTGACCAACCTCCCGGTGACACACTCACCACCTGCT 900
Db 1803 GCACCTCTGACGAGGACACTGAGTCACTGACCAACCTCCCGGTGACACACTCACCACCTGCT 1862
QY 901 CGCCCTCATTTCCCGGGTAAGTCACTGACCAACCTCCCGGTGACACACTCACCACCTGCT 960
Db 1863 CGCCCTCATTTCCCGGGTAAGTCACTGACCAACCTCCCGGTGACACACTCACCACCTGCT 1922
QY 961 GCACCCCGGGCTTCTGCGGCCCTCGGCCAGCTGCTGAGGAAGCCGAGGCCAGTGA 1020
Db 1923 GCACCCCGGGCTTCTGCGGCCCTCGGCCAGCTGCTGAGGAAGCCGAGGCCAGTGA 1082
QY 1021 GCGTTCACGAGGCCCTCAGCTCGCCCTGCACTCTCCGAGGGTCCCGGCCACCTCGGCCG 1080
Db 1983 GCGTTCACGAGGCCCTCAGCTCGCCCTGCACTCTCCGAGGGTCCCGGCCACCTCGGCCG 2042
QY 1081 CCAGGGCTGCGTCCCGGACAGCTCATCCGACAGCTGTTCTACCAACAGGCGGGTGA 1140
Db 2043 CCAGGGCTGCGTCCCGGACAGCTCATCCGACAGCTGTTCTACCAACAGGCGGGTGA 2102
QY 1141 GAGCCCTGAGCA CGGGGCTTCCGCGGCCGAGAGCGGCGCACCCGCGCCCGCGCGCGCA 1200
Db 2103 GAGCCCTGAGCA CGGGGCTTCCGCGGCCGAGAGCGGCGCACCCGCGCCCGCGCGCGCA 2162
QY 1201 CCTGCTGAAGGGCGGGTACATTGCGGGGCGCTGCCAGGCTGCGGAGCCACTGATGGA 1260
Db 2163 CCTGCTGAAGGGCGGGTACATTGCGGGGCGCTGCCAGGCTGCGGAGCCACTGATGGA 2222
QY 1261 GCACCGGGTGTGAGGAGGAGGCGCGCAGGGAGGAGCAGGCCACCTCTCGGCCAAGC 1320
Db 2223 GCACCGGGTGTGAGGAGGAGGCGCGCAGGGAGGAGCAGGCCACCTCTCGGCCAAGC 2282
QY 1321 CCCCTCATTCGAGACTGCCCCCTCGGCTGCTGCTGCGCACCCACTTGGCCCCCTGCGCA 1380
Db 2283 CCCCTCATTCGAGACTGCCCCCTCGGCTGCTGCTGCGCACCCACTTGGCCCCCTGCGCA 2342
QY 1381 CAGCCACTCCCTGGAACATGACTCTCGAGACACCCCGCGGCCCTCTCGAGGGCTGCGG 1440
Db 2343 CAGCCACTCCCTGGAACATGACTCTCGAGACACCCCGCGGCCCTCTCGAGGGCTGCGG 2402
QY 1441 TGAGGACAGGCACTGCTTCAGCCCTCGCGGGGCGCCCTATCAGGGACATGGGGCA 1500
Db 2403 TGAGGACAGGCACTGCTTCAGCCCTCGCGGGGCGCCCTATCAGGGACATGGGGCA 2462
QY 1501 CCCTCAGGGCTCCAGCAGCTTCCATCCACTGCTGGCCACCCAGGCACTGCTCAGCCAGA 1560
Db 2463 CCCTCAGGGCTCCAGCAGCTTCCATCCACTGCTGGCCACCCAGGCACTGCTCAGCCAGA 2522
QY 1561 GAGGCCATCCCGGACAGCCCTTGGGGGACGCGCCCTTTCTGCGCACCCCAAGCAGGG 1620
Db 2523 GAGGCCATCCCGGACAGCCCTTGGGGGACGCGCCCTTTCTGCGCACCCCAAGCAGGG 2582
QY 1621 TTCTGCCCCCAGAGGGCTCAGCCCCCAGCAGAGTTGCCCTCCTGCGCTC 1680
Db 2583 TTCTGCCCCCAGAGGGCTCAGCCCCCAGCAGAGTTGCCCTCCTGCGCTC 2642
QY 1681 CTTCCCTCCAGGATCTTTGCAAGAGGCCCTTTAGTACCTTCAAGCCCTCTTTGGGACA 1740
Db 2643 CTTCCCTCCAGGATCTTTGCAAGAGGCCCTTTAGTACCTTCAAGCCCTCTTTGGGACA 2702

QY 1741 GCCCCAGGACCCCTGCTGCTCAAGCAAGCCCTCCCTATTGGACTCTAAGATGGGGCC 1800
Db 2703 GCCCCAGGACCCCTGCTGCTCAAGCAAGCCCTCCCTATTGGACTCTAAGATGGGGCC 2762
QY 1801 TGAGAGACATCTCTCTCTGGGAGGCCAAAACCCGGGCCCTCGCAGTTCCCGAGGGTCAGC 1860
Db 2763 TGAGAGACATCTCTCTCTGGGAGGCCAAAACCCGGGCCCTCGCAGTTCCCGAGGGTCAGC 2822
QY 1861 CTCCAGGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCCCAAGTGGGAC 1920
Db 2823 CTCCAGGCGAGCTCTTCCCAAGTGAAGTCCCTCAGGGTGGGCTCTCCCAAGTGGGAC 2882
QY 1921 AGACCTTGCCCTCCCTTGATGCGAGGGCTGACCCAGGAGCTGAGGATCTGTCCGA 1980
Db 2883 AGACCTTGCCCTCCCTTGATGCGAGGGCTGACCCAGGAGCTGAGGATCTGTCCGA 2942
QY 1981 CTCCACACCCACCTTGACGCGGCCCTCAGGAACAGCGACCAATGCGCAAGTTCTCCCTGG 2040
Db 2943 CTCCACACCCACCTTGACGCGGCCCTCAGGAACAGCTGACCAATGCGCAAGTTCTCCCTGG 3002
QY 2041 TGCTCGCGGGGTACGACAGGCTGGCTTATGGCACTTTGCTTGGTGAGATGC 2100
Db 3003 TGCTCGCGGGGTACGACAGGCTGGCTTATGGCACTTTGCTTGGTGAGATGC 3062
QY 2101 AGGGGCACTGCTGGGCGAGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGCCAGTC 2160
Db 3063 AGGGGCACTGCTGGGCGAGGGCCCATGTGGGCCAGGATAGCTGGGCTGTGCCAGTC 3122
QY 2161 GGAGGAGGAGAGCAGAGGAGGCGCAGGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 2220
Db 3123 GGAGGAGGAGAGCAGAGGAGGCGCAGGGCTGAGTCCAGTCGAGGAGCAGCAGGAGGC 3182
QY 2221 CAGGGCTGAGAGGCCACTGCCCCAGGTCAGTGCAGAGGCTGTGCTGAGTTCGGCAGGGC 2280
Db 3183 CAGGGCTGAGAGGCCACTGCCCCAGGTCAGTGCAGAGGCTGTGCTGAGTTCGGCAGGGC 3242
QY 2281 TCCACACAGAGGCTCTCCAGAGCCACCCATGGGAGGACATCGGGCAGGTCCTCCCTGT 2340
Db 3243 TCCACACAGAGGCTCTCCAGAGCCACCCATGGGAGGACATCGGGCAGGTCCTCCCTGT 3302
QY 2341 GCAGATCCGGGACCTGTGAGTGTGAGGAGGCGGCCGACAAATATCCCTGGACATTTTC 2400
Db 3303 GCAGATCCGGGACCTGTGAGTGTGAGGAGGCGGCCGACAAATATCCCTGGACATTTTC 3362
QY 2401 CGAGGTGGAACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 2460
Db 3363 CGAGGTGGAACCCCGCTTACCTCAACCTCTCAGACCTGTACGATATCAAGTACCTCCCAT 3422
QY 2461 CGAGTTTATGATCTTCAAGGAAGTCCCAAGTCCGCTCAGCCAGGCGCCCTCCCCCAT 2520
Db 3423 CGAGTTTATGATCTTCAAGGAAGTCCCAAGTCCGCTCAGCCAGGCGCCCTCCCCCAT 3482
QY 2521 GGCTGAGGAGGAGCTGGCCGAGTTCCCGGAGCCACCGTGGGCCCTGGCCAGGTGAACCTGGG 2580
Db 3483 GGCTGAGGAGGAGCTGGCCGAGTTCCCGGAGCCACCGTGGGCCCTGGCCAGGTGAACCTGGG 3542
QY 2581 CCCCCACGCAAGCTGAGATCAGAGGAGTCAAGGAGTGTGACGCGCTGTCTGGGACAGA 2640
Db 3543 CCCCCACGCAAGCTGAGATCAGAGGAGTCAAGGAGTGTGACGCGCTGTCTGGGACAGA 3602
QY 2641 GGCTGCGGTGGGAGGAGGCGAGTGGTCTCGCCCTCAGCAGACCTCTTCCACTTCCC 2700
Db 3603 GGCTGCGGTGGGAGGAGGCGAGTGGTCTCGCCCTCAGCAGACCTCTTCCACTTCCC 3662
QY 2701 TGGGAGGACACTGCGCTGATGAGCTGACAGCTGGGGCTGGGCTGAGAGAGTGAAGGC 2760
Db 3663 TGGGAGGACACTGCGCTGAGCAGGCTGACAGCTGGGGCTGGGCTGAGAGAGTGAAGGC 3722
QY 2761 CTCCGTGGAGCACAATCTCCCGGATCTTAAGGGCAGGCCGGAAGGTCTGGAGAGAGGG 2820
Db 3723 CTCCGTGGAGCACAATCTCCCGGATCTTAAGGGCAGGCCGGAAGGTCTGGAGAGAGGG 3782
QY 2821 GCCCCCAGGAGAGGAGCCAGGCTTGTCTTCCCGGCTCTCAGGTCTGAGAGAGCTGGGA 2880

D _b	3783	GCCCCCAGGAAGAAGCCGCGCTTGCTTCCTCCGGCTCTCAGGTCTGAAGAGCTGGGA	3842
Q _y	2881	CCGAGGCCGACATTTCTTAAGGGAGCTTCTCAGATGAGACTGTGGTCTCTGGCCGAGTCAGT	2940
D _b	3843	CCGAGGCCGACATTTCTTAAGGGAGCTTCTCAGATGAGACTGTGGTCTCTGGCCGAGTCAGT	3902
Q _y	2941	GACACTGGCCTGCCNGSTGTCAAGCCGACAGCTGCCAGGCCACCTGGAGCAAAGACGG	3000
D _b	3903	GACACTGGCCTGCCAGSTGTCAAGCCGACAGCTGCCAGGCCACCTGGAGCAAAGACGG	3962
Q _y	3001	A 3001	
D _b	3963	A 3963	

Search completed: March 21, 2005, 16:08:22
Job time : 1893.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:24:10 ; Search time 4515.62 Seconds
(without alignments)
10741.310 Million cell updates/sec

Title: US-10-077-130-4_COPY_10500_11500
Perfect score: 1001
Sequence: 1 gccctgtgagtgaggaa.....ttctagaagatgtgaaaaa 1001

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :		GenEmbl.*	
		1: gb_ba.*	
		2: gb_hgt.*	
		3: gb_in.*	
		4: gb_on.*	
		5: gb_ov.*	
		6: gb_pat.*	
		7: gb_ph.*	
		8: gb_pl.*	
		9: gb_pr.*	
		10: gb_ro.*	
		11: gb_ats.*	
		12: gb_av.*	
		13: gb_un.*	
		14: gb_vi.*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	20435	9 HSAJ2535	AJ002535 Homo sapi
2	997.8	99.7	4789	9 AB046776	AB046776 Homo sapi
3	714	71.3	18524	6 CQ730656	CQ730656 Sequence
4	343.4	34.3	352	6 AX069754	AX069754 Sequence
5	335.2	33.5	34667	9 HSA314905	AJ114905 Homo sapi
6	335.2	33.5	135964	9 AL3353593	AJ1353593 Human DNA
7	321.6	32.1	164766	2 AC028657	AC028657 Homo sapi
8	321.6	32.1	174612	2 AC023889	AC023889 Homo sapi
9	278.6	27.8	164766	2 AC028657	AC028657 Homo sapi
10	276	27.6	260998	2 AC099089	AC099089 Rattus no
11	267.2	26.7	211829	10 AL645854	AL645854 Mouse DNA
12	238.8	23.9	2155	6 AX463538	AX463538 Sequence
13	193.8	19.4	2295	9 HSM800639	AL080128 Homo sapi
14	185	18.5	621	11 BV014682	BV014682 S212P6028
15	107.4	10.7	2534	6 AX834822	AX834822 Sequence
16	107.4	10.7	2534	9 AK097489	AK097489 Homo sapi
17	105.8	10.6	2761	9 BC043916	BC043916 Homo sapi
18	86	8.6	1005	6 CQ445545	CQ445545 Sequence
19	82.6	8.3	168144	2 CR792456	CR792456 Danio rer

20	74.8	7.5	98613	2	AC139630	AC139630 Takifugu
c 21	71.4	7.1	178817	2	CR394562	CR394562 Danio rer
c 22	67.6	6.8	7218	6	I66494	I66494 Sequence 14
c 23	63.8	6.4	180872	5	AL844185	AL844185 Zebrafish
c 24	63.4	6.3	125020	9	AF429315	AF429315 Homo sapi
25	62	6.2	3935	6	AX833410	AX833410 Sequence
26	62	6.2	3935	9	AK095113	AK095113 Homo sapi
27	60	6.0	60	6	CQ543881	CQ543881 Sequence
28	59.4	5.9	228100	2	AC097091	AC097091 Rattus no
29	59	5.9	89623	6	CQ730839	CQ730839 Sequence
30	59	5.9	93801	6	BD186121	BD186121 Preventio
31	59	5.9	93805	6	CQ730837	CQ730837 Sequence
32	59	5.9	103053	6	CQ848088	CQ848088 Sequence
33	58	5.8	303091	2	AC084799	AC084799 Mus muscu
34	57.8	5.8	5378	6	CQ875354	CQ875354 Sequence
35	57.8	5.8	5783	6	CQ715191	CQ715191 Sequence
36	57.8	5.8	185079	9	BS000240	BS000240 Pan trogl
37	57.4	5.7	31595	6	AX780060	AX780060 Sequence
38	57.4	5.7	81940	9	HSTIIN2B	X90568 H. sapiens m
39	57.2	5.7	6602	4	AY136512	AY136512 Canis fam
40	57.2	5.7	233611	2	AC112465	AC112465 Rattus no
41	56.8	5.7	53424	3	AB055861	AB055861 Procamb
42	56.4	5.6	5382	9	AB014557	AB014557 Homo sapi
43	54.8	5.5	298166	2	AC087563	AC087563 Homo sapi
44	53.6	5.4	3316	9	BC007201	BC007201 Homo sapi
45	52.8	5.3	205691	2	AC087227	AC087227 Mus muscu

ALIGNMENTS

RESULT 1	HSAJ2535	Homo sapiens mRNA for obscurin (OBSCN gene).	20435 bp	linear	PRI 14-SEP-2001
LOCUS	OBSCN				
DEFINITION	OBSCN				
ACCESSION	AJ002535				
VERSION	AJ002535.1	GI:15026973			
KEYWORDS	OBSCN gene; obscurin.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1	Young, P., Ehler, E. and Gautel, M.			
AUTHORS	1	Young, P., Ehler, E. and Gautel, M.			
TITLE	1	Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly			
JOURNAL	1	J. Cell Biol. 154 (1), 123-136 (2001)			
MEDLINE	21342081				
PUBMED	11448995				
REFERENCE	2	(bases 1 to 20435)			
AUTHORS	Gautel, M.S.				
TITLE	Direct Submission				
JOURNAL	Submitted (31-OCT-1997)	Gautel M.S., Structural Biology Division, European Molecular Biology Laboratory, Meierhofstr. 1, Heidelberg, 69117, GERMANY			
REMARK	Revised by author 20-JUL-2001				
FEATURES	Location/Qualifiers				
source	1. .20435				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/map="1q42.1"				
	/cell_type="cardiomyocyte"				
	/tissue_type="cardiac"				
	/dev_stage="adult"				
	1. .20435				
	/gene="OBSCN"				
	72. .19934				
	/gene="OBSCN"				
	/functions="titin binding"				
	/codon_start=1				
	/evidence=experimental				

/product="obscurin"
/protein_id="CAC44768.1"
/db_xref="GI:15026974"
/db_xref="GOA:O96AA2"
/db_xref="UniProt/T/EMBL/O96AA2"
/translation="MDQPOFGSAPRLTRPKAFVSVSGKDATLSQIVGNPTFPQVSWE
KQCPQAFGRFLAQDGLYRLTLTLDALDSGGQVVCRAARNAIEAFAAAGLVQDAE
AACQAPQAFGLRLPTSVIRREGSEATFRCRVGSGPVPVSWSKDGRGLGEPDPRVVR
BELGEASALIRBAAPRDGGTYVEAENPLGAASAAALVVDSDAADTASRPGTSTA
LLAHLORREARMAEAGAPASPPSTGTCTCTTEGKAARLSCTVCEPEKPTVMKQDG
LVTEGRHVVEDEAENFKVLKFCQSDRGLYCTCTASNLVGQTVSSVLVVREPAVP
FKRLQDLVREKESATFCEVQFSTERAMKEETRLWASAKYIEEGEERRLTVR
NVSADDAVICTETEGSTVAELAQGNLRLKPKRTAARVGDGMTCEVLAIPVPG
VHMLRQNEVAGGVAISAEGRHTLTISQCLLEDQGVAFMGDCOTSTRFCVSP
RKPPLQPPVDFVKARMSVILSWSPPHGPERPVTIDGLYVEKKLCTYTWIRCHEA
EWVATPTEADVAESGPOPRVSALEFGOSPYLEPFGTVHLAPKLAVRPLKAVQA
VGGVEFTSDVLTASCEMFLDQALKASSVYELHCDRTHTLTIRVPASLHQAQ
KPVANGILESSIRMEVRAAPGLTANKPAAARAEVLAHEAQAALLAEQDAATWML
KQRTLSFGPKTYEQASGRVLLVRDARDAGLYECVSRGGRIAYQLSGGLARL
KDMAGSCVDAVAGGPAQFECETSEAHVYVHMDGMLHSGRERFLQEDVGTGRLV
AATVTRQDEGTYSCRVGSDVDFRLRVSEPKVFAKEQLARRKLQAEAGASATLCEV
LDVKEPKVFAQDVAAHSEVQAEAGANATLCEVAAQAEVWYKDKKLSLSLVHV
EAKGCRRLVQAQAKTIDAGDYSCEARQVRSGFRUHIITEPKOMFAKQSVINEVQAE
GASAMLSCEVAQAQKTVIYKDKKLSKSSKVMVEVKGCTRLVLPQAKADAGEYC
EAGGQRVSFHLHIITEPKVFAKEQSVHNEVQAEAGTATMLSCVQAQPTQVETWKDGK
KLSSSKVMVEKGTCTRLVQAQKADAGEYSCVAGQVRVSGFOLHIITEPKAVFAKEQ
LVHNEVTEAGASATLCEVQAQAEVTVIYKDKKLSKSSKVRTEAAGCMRLVVQAQ
GOADAGEYCEAGGQRLSFHLDSPEKAVFAKEOLAHKRVQAEAGATLSCVLAQV
TEVTVYKDKKLSKSSKVMVEVKGCTRLVQAQKADAGEYSCVAGQVRVSGFOLHI
EPKVPFAKEQVHREVQAQASATLCEVQAQAEVWYKDKKLSKSSKVMVEVKGCTRL
CTRLVQAQKADAGEYSCVAGQVRVSGFOLHIITEPKOMFAKQSVINEVQAEAGT
TSLCEVAQAEVTVIYKDKKLSKSSKVMVEVKGCTRLVQAQKADAGEYSCVAGQ
QRLSFHLVABEPKVFVFAKEQPAHREVQAEAGSATLCEVQAQAEVTVIYKDKKLS
SSKVRVAGCTRLVQAQKADAGEYSCVAGQVRVSGFOLHIITEPKAVFAKEQ
PLVKEHEDIILTATLTPSAATVWLKQDVEIRSKRHTFASQDTHLTITVGAQVIL
DSALYSCVGAEGQDPVQVREVAACFRLLEPVCGELGTITVLACLSFACAEVWVR
CQNTQPRVKGAFQVAGFVRSVLVLGLRAEDAGEYCEVSRDHTSQAQVSVPRVVK
FMSGLTVABEGEATPQCVSPSDVAVVMDGALLQPSSEKFAISQSGASHLSTIS
DLVLEDAQTITVABEGASSAALVREAPVLFKKLBPQTVSEERSVTLSEVLETRWPM
ELRWTRNATLAPGNVIEHAEGARHLVLHNVGFADRGFGCEPTDDTKQAQKLTVM
RQTLVRGLQAEAREQOATVMEVQLSHADVDSGWRDLRGFOQGPCTCHLAVRGMHT
LTLGLRPEDSGLMFAEGVHTSARLVTLVPSFSRPLQDVVTEKEKVTIQAQVIL
PNVDVWLKQDGLVLRAGKTMALAAQACASLTITVYCFADQVQVCDHADESSASV
KQVQRTYLTIRVLAEDAGEIQVNAENSAQLRVKELPVLTVLRDLKIAMKXHR
GVLCQVSRASQVFGQELQPKYELSDGLYRKLISDVAEDEDYTCDAQ
DNVTSQAFVVEEQITIVRGQVTVMEPAPAMPCECTSPSVRPDKMLGKTLQAG
VWGLEQEGTVHRLMLRRTCTMTGTPVHTVKGSRSSARLVVSDIPVILTRPLEPKTG
RELQSVLSCDFRPAKAVQYKDDTPLSPSEKFKMSLEGQMAELRLRLMPADAGVY
RCQAGSAHSSTEVTEAREVTVGLQDAEATEGWSAFSCSLSHDEHEVSWLGNMP
LYNDSFHEISHKRRHTLVLSIQRADAGIVRASSALKVSTSLARLEVRKVEVFLKALD
DLSABERGTLALQCEVSDPEAHVVRKDGVLQSPDKYDFLHTAGTRGLVHVDVSPED
AGLYTCHVGBEETARVRVHDHVLGHTIKRLKTMVELEGESCEVLSHESAGDPAMW
TVGKTVGSSSRFOGTRQGRKYILVREAAAPSDAGEVVSFVSGVLTSKASILVREPA
I IKPLEQWVAPGDVRLCELSRAGTPVHMLKDRKAIKRSQKYDVVCEGTAMLVIR
GASLKDAGETVCEVASKSTASLHVEKANCFTBELNLQVBEKANTVCKTEHPAA
TVVWKGLELRAEGKQPSQBLTLRLTISALEKADSDTYTCDIQOASRAQLVYQG
RVHKLIDLEDDVDVQESSATFRCRISANIEFVHPLFDKTLPLHMLNEIDAQPGVY
HVLTLRLQALQKSGTYIFEAGDQASAAALVTEKPSVFSRLTDATITEGDELTVCE
TSTCDIIPMCWTKDKTLRGARCOLSHEGHAOLITGATLQDSQKRYCEAGACACSS
IVRVHAPVRFQALKDLVLEGAATLRCVLSVAAPVWKYGNVLRPGDKYSLRQ
EGAMLFLAVNRLPQDSGRYSCSFGDTTSATLTITVATLPAQFICKLRNEATEGATAT
LRCELSKTAPEVWRKSGSETLRDGDYCLRDQGWCELIQGLAMVDAEYSCVGBER
TQSASLTIPMPAHLIIGLRHQSIEBAGATLRLCELSKAPVPEWRKGLRSDRHS
RQDQAVCELOQICGLADAGEYSCVCEERTSATLTVPKALPAKFTGLRNEAVEGAT
AMLWCELSKVPVPEWRKGPENLRDGDYILRQEGTRCELIQICGLMADAGEYLCVQGY
ERTSATLTIRALPARFIEDVKNQAREGATVLOCELRCELIQICGLMADAGEYLCVQGY
SLRQDGTKECLOIRGLAMADTGEYSCVCGOERTSMLTVRALPIKFTTEGLRNEATEG
ATLVRCELSKMAPEVWKGHETLRDGDYLRDQDHSLRQDGCARCELQIRGLVADAGEYLC
KERTSAMLTVRAMPKFTTEGLRNEATEGATLWCELSKAPVPEWRKGLRSDRHS
RHSRLQDGSCELIQIRGLAMADTGEYSCVCGOERTSATLTIRALPARFIEDVKNQARE
EGATAVLQCELSKAPVPEWRKGLRSDRHSRQDGTKECLOIRGLVADAGEYSC
VCGOERTSATLTIRAPQVPRFELQSLQAEBSATLQCELSKAPVPEWRKGLRSDRHS

ANRRERPLRQCTAELVLQDLQREDTCEYTCGSOATSLTLYTAAVPRFLRLQHQ
EVDGEGTHLCCELSRAGASVWRKGLQLFPCKAKVQVQGAALRLVGRVQEDAG
DYTCDTGHTQSMASVVRPFRPKFKTLQSLQEDTGDIARLQSDAESAGVQWMLK
EGLVHAGPKYEMRSQATRELLHQLQEAQDTGEYACVTGQKTAASLRTVEPTIV
RGLVDAEDTADVEDFSCVSRAGATGVQCLQPLQSQNEVTEVAVRQDRIHLRLK
RGTVPEDAGTVSFHLGNHASSAQLTVRAPEVITILEPLQVOLSEGQDASFCRLSRASG
QEARWAGVPLQANEMNDITVBOGTLHLTLHKVTLLEDAGTVSFHVGTCSSEALQKV
TAKNTVVRGLVNEALEGGALPECOLSQEVAHAHTWLLDDEPVTSENAEVVFFENG
LRHLLLLKNLRPQDSCKRTVFLAGMDVSAFLTVRGRLELLEPLKNAARVAGAART
CTUSEVAVGEASWYINGAAVQDDSDMTVTADGSHQALLRSLAQPHAGEVTFACDR
AVASARLTVLGLPDPEDAEVVAHSTVLSWAAQPMSCGGGLCGVRVVEKATQ
PSAPESQVAAAGDVSLELEVAEAGEVYHKGWERIOPGCRPEVVSQGRQMLYKIG
FLAECRMSRPTLDSISELPEEDGRSQRLEAEVAPDSEGSTADELARTGADL
SHTSDDSEFAGTPTSLVYLKAGRFGTSPASKVGAAPASVYKQOQBPPLAARPP
LDGLSTKDLGDSMDKAAVKIQAFKGYKVRKEMQOEGPMFTEGDTAEQVGDALR
LECVASAKADVRLWKDQVGLTDRHHHIDQLGDGTCSLLIAGLRADAGCVTCVS
NKFQGVTHSACVVVSGSEARSSSGGLDDAARRARLRLHFRKSPAEVSDLEP
LSADEGPAPEEPADMTVREDHFICIRPEALTEARQAVTRQEMPAILGIVELK
VEQPRVEMCISKETPAVPPPEPLSLTSDAAPVFLTELQNEVQDGVPSFDCV
VTQPMPSVRWFQKDLLEDHDMINEDQGGHQLIITAVPADMGVYRCLAENSG
VYSTKABLVRDLTSTDYDAADATESSYFSAQYLSREGEQTESTDEGQLPVVE
ELRLQVAPGTRIAKQLKVGYPARLVYFKQGPPLTASAHIMQDKTLHTLEIIS
VTRDGGQYAAVTSNAGAAVSARLLVRGPDEPEKPSADVHEQLVPRMLERFPTK
KVKGSSITFSVVKGRPVTVHVLREAEVGLVTPDTPGYTVASSAQHSLVLLD
VGRHOGTCTVTSNAGQALCSASHLSGLPKVREQKKEALISTFLOGTQALISA
QGFOTASFADLQGRKEEPLAKAHLGLAEVTEEFLOKLTQITEMWSAKITQA
KUGPGGDSDESKTSPSRHGRSPSSIQESSSESDGARGVIFDIYVVTADYL
PLGAEDAITLRREGYVFLDAARLPLRVTKTSPSRQVSWSPAYLDRRLKLS
EWGAABFPFGAEVSEDEYKARLSVIOELLSQEAQVFELOQLSHLQHLRCPH
VPLAVAGQAVIFRNVDRIDGRHSSFLQELQCDTDDVAMCFIKNOAEPLOYLEFL
GRVQSVVSTALOFYKYKAEALLAGDSQPPLQLHYLEQVVERVQYQALLK
ELIRKARNQNCALLQYAVVSALPQAEKHLVSLMENTPGTLQAGEPPIRQHP
IWEVGARMPKWNHVRFLRNLVICKPRRSDRTVSVYFNMMKLSLIDND
QVEDDRAEVEVQEREDSVRYLLQARTAIKSSWVKEICGIQORLALPWRPDPFE
ELADCTAEISETVLAARVGTGTPKPVISYKDKGAVQVDPHILLIDPDGSCALILDS
LTVGQYKQWCFAAAGNCSTKILVQVPPRVFNKVLASPFVEGEDAFLRTIGA
PVPOTRWYKQKALLTNGKFTLSRPSGLLVILRAASKDLGJECLELVNRLGASR
ASAEILQSPMLQAOQCHREOLVAARETEGETKVPKTVIIBETITVVKSPGRQ
RSPSPSPSPSCASPLRPGILLAPDILLGAGQPRRPEAPPOKVPVPLYVTEA
EASHPALGSLGSPKSPVEVEITVVRKMGQGVSPVTPRSPSGHFLTPGATP
GADPNNNNKLLAQEAQWAGTMVREPLVFRVDARSDVMAASGMGSLSEEGTM
EABEGEEDGDAFTVEESQDTHSLGDRDKPLITHNGRMLTLADLEDVPGEGEYFHC
CGPGCAPDPPCEVSVIOREIPEPTQOPVLLSVGHALGPRGLGLPEPRGASPP
GFQVRSLEGTSLLEAPAPVGSQVPSQVFCFTRIRRSADSSQSSFTTSLTQTVNFQ
TVGEVTVLHICFDROGDENAQP

ORIGIN

		Query Match	100.0%	Score 1001;	DB 9;	Length 20435;
		Best Local Similarity	100.0%;	Pred. No. 1.9e-203;		
		Matches 1001;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GCCTCTGTGAGTGGAGGAGGGCCGAGAACCTCAGAGATGGGACAGATACATCTCTG	60			
Db	10500	GCCTCTGTGAGTGGAGGAGGGCCGAGAACCTCAGAGATGGGACAGATACATCTCTG	10559			
Qy	61	AGGCAGAGGGGACACAGGTGTGAGTGCAGATCTGTGGCTGGCCATGCGGACGCCGGG	120			
Db	10560	AGGCAGAGGGGACACAGGTGTGAGTGCAGATCTGTGGCTGGCCATGCGGACGCCGGG	10619			
Qy	121	GAGTACTTGTGTGTGTCGGGACGAGAGGACCTCAGCCACGCTCACCATCAGGCTCTG	180			
Db	10620	GAGTACTTGTGTGTGTCGGGACGAGAGGACCTCAGCCACGCTCACCATCAGGCTCTG	10679			
Qy	181	CCTGCCAGGTCATAGAGATGTGAAAAACAGGAGGCGCAGAGAGGGGCCACGGCTGTG	240			
Db	10680	CCTGCCAGGTCATAGAGATGTGAAAAACAGGAGGCGCAGAGAGGGGCCACGGCTGTG	10739			
Qy	241	CTGCAGTGTGAGCTGAAACAGTGCAGCCCTCTGTGAGTGGAGAAAGGGGTCTGAGACCTC	300			
Db	10740	CTGCAGTGTGAGCTGAAACAGTGCAGCCCTCTGTGAGTGGAGAAAGGGGTCTGAGACCTC	10799			

QY 301 AGAGATGGGACAGATACAGCTGAGGACGGGACCTAAATGTGAGCTGCAGATTCGT 360
|||||
Db 10800 AGAGATGGGACAGATACAGCTGAGGACGGGACCTAAATGTGAGCTGCAGATTCGT 10859
QY 361 GGCCTGGCCATGGCAGACACTGGGAGTACTCGTGTGCGGACGAGAGACCTCG 420
Db 10860 GGCCTGGCCATGGCAGACACTGGGAGTACTCGTGTGCGGACGAGAGACCTCG 10919
QY 421 GCTATGCTCAGCGTCAGGCTTACCCATCAAGTTACAGAGGGTCTGAGAAAGAG 480
Db 10920 GCTATGCTCAGCGTCAGGCTTACCCATCAAGTTACAGAGGGTCTGAGAAAGAG 10979
QY 481 GCCACAGAGGGGCAACAGCGCTGCTGGGTGAGCTGAGCAAGTGCCTGGTGGAG 540
Db 10980 GCCACAGAGGGGCAACAGCGCTGCTGGGTGAGCTGAGCAAGTGCCTGGTGGAG 11039
QY 541 TGGTGGAGGGGCGATGAGACCTCTAGAGATGGAGACAGACAGCTGAGGACGAGCGG 600
Db 11040 TGGTGGAGGGGCGATGAGACCTCTAGAGATGGAGACAGACAGCTGAGGACGAGCGG 11099
QY 601 GCCAGGTGAGCTGAGCTGCGGGCTCTGGGACAGAGAGCTGGGGAGTACTGTGC 660
Db 11100 GCCAGGTGAGCTGAGCTGCGGGCTCTGGGACAGAGAGCTGGGGAGTACTGTGC 11159
QY 661 ATGTGCGGGAAGGAGGACCTCAGCCATGCTCACCGTCAGGGCCATCGCTTCCAAGTTC 720
Db 11160 ATGTGCGGGAAGGAGGACCTCAGCCATGCTCACCGTCAGGGCCATCGCTTCCAAGTTC 11219
QY 721 ATAGAGGCTCTGAGGAATGAAGAGGCCACAGAGGGGACAGCGGACCGCTGTGTGAG 780
Db 11220 ATAGAGGCTCTGAGGAATGAAGAGGCCACAGAGGGGACAGCGGACCGCTGTGTGAG 11279
QY 781 CTGAGCAAGGGCGACCGCTGAGAGTGGAGAGGGGACATGAGACCTCAGAGATGGGAC 840
Db 11280 CTGAGCAAGGGCGACCGCTGAGAGTGGAGAGGGGACATGAGACCTCAGAGATGGGAC 11339
QY 841 AGACACAGCTGAGGACGAGCGGTCCAGGTGTGAGCTGAGCTCGAGTCCGTGGCTGTG 900
Db 11340 AGACACAGCTGAGGACGAGCGGTCCAGGTGTGAGCTGAGCTCGAGTCCGTGGCTGTG 11399
QY 901 GTGATGCGGGGAGTACTCGTGTGTGTGCGGGACGAGAGACCTCAGCCACACTCACT 960
Db 11400 GTGATGCGGGGAGTACTCGTGTGTGTGCGGGACGAGAGACCTCAGCCACACTCACT 11459
QY 961 GTCAGGCGCTGCTGCTGCGAGATTCATAGAAGTGTGAAAA 1001
Db 11460 GTCAGGCGCTGCTGCTGCGAGATTCATAGAAGTGTGAAAA 11500

RESULT 2
AB046776
LOCUS AB046776 Homo sapiens mRNA for KIAA1556 protein, partial cds. PRI 22-FEB-2001
DEFINITION
ACCESSION AB046776
VERSION AB046776.1 GI:10047176
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
REFERENCE
AUTHORS Negase,T., Kikuno,R., Nakayama,M., Hirose,M. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XVII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
JOURNAL DNA Res. 7 (4), 273-281 (2000)
MEDLINE 20450683
PUBMED 10997877
REFERENCE
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba

292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914)
FEATURES
source
location/Qualifiers
1. 4789
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="flh18619"
/note="vector:pBluescriptII SK plus"
1. 4789
/gene="KIAA1556"
1. 4789
/gene="KIAA1556"
/note="Start codon is not identified."
/codon_start=3
/product="KIAA1556 protein"
/protein_id="BAB13382.1"
/db_xref="GI:10047177"
/translation="RAVRVHDLHVGITKRLKTNVLEGGSCSPCVLSHESASDPM
WTGKTVGSSRRFQATRGKRIYLVREAPSDAGEVVFVSRGLTSCAGSLIVRERPA
AIIKPLEDQWVAPGEDVELRCELSRAGTPVHMLKDKRAIKRSQYKDVVCEKTMALVI
RGSLKDAGYTCVEASKSTASLHVEEKANCFTEELTNLQVEEKGTAFTCTKTHPA
ATVWRKGLLELRASGHQSQEGLRLTLTSALEKADSDTYTCDIGQAQRAQLVQ
GRRVHIIEDLEDVDQSSSATPRCISPNAYEPVHFLDKTPIHANELNEDIDAPGG
YHVLTLRLALKDSGTIYFAGDQASALRLVTEKPSVFSBELTDATITEGEDTLVC
ETSDCDIPVMTKDGKTLRSGACQLSHEHRAQLITGATLQDSGRYKCSAGACSS
SIVRVHAPVRFQALDLVLEGGATLRCVLSVAAPVKNVGNVLRPGDKYSLR
QEGMALVLRNLRPDSGRYSCSGQDTTSATLVTALPAQFIKLRNKEATGEGATA
TLRCELSKAAPEVWRKGETLRDGDYCLRDQGANCELOIRGLAMDAEYSCVCGEE
RTSASLTIRWPAHFIGRLRHOESIEGATATLRCELSKAAPEVWRKGRESLRDGRHS
LRQDGVCELOICGLAVADAGEYSCVCGEERTSATLTVKALPAKFTGLRNEAVEGA
TAMUWCELSKAAPEVWRKGPENLRDGRYILRQEGTCELOICGLAMDAEYLCVCG
QERTSATLTIRALPARFIEDVKNQAREGATVLOCELSNAAPEVWRKGETLRDGR
YSLRQDGTCELOIRGLAMDAEYSCVCGEERTSATLTIRALPIKFTGLRNEATE
GATLVRCELSKAAPEVWRKHETLRDGRHSRLQDGCARCELOIRGLVAEDAGELCM
CGKERTSAMLTVRAMPKFTGLRNEATEGDTATLMCELSKAAPEVWRKHETLRDGR
REGHSLRQDGRCELOIRGLAVADAGEYSCVCGEERTSATLTIRALPARFIEDVKNQEA
RHSALVRCELSKAAPEVWRKGETLRDGRYILRQEGTCELOICGLAMDAEYLCVCG
QERTSATLTIRALPARPTQDLKASEGATATLQCELSKAAPEVWRKGETLRDGR
DGRYSILRQDGTCELOIRGLAMDAEYSCVCGEERTSATLTIRALPARFIEGLRNE
EMEGATATLQCELSKAAPEVWRKLEALRDGKYSILRQDGCARCELOIRGLAMADNGV
YSCVCGEERTSATLTIRALPARFIEDVKNQAREGATVLOCELSKAAPEVWRKGPNT
LKDDGRYSILRQDGTCELOIRGLAVADAGEYSCVCGEERTSATLTIRALPARFIEDVR
NEATEGATATLQCELSKAAPEVWRKGETLRDGRYSLRQDGTCELOIRGLAVEDT
GHELCVCGEERTSATLTIRALPARFIEDVKNQAREGATATLHCELSKAAPEVWRKGP
ETLRDGRHSILRQENLRNLPNGGGGSELGSCSPAWTAKLDFIS"

ORIGIN

Query Match 99.7%; Score 997.8; DB 9; Length 4789;
Best Local Similarity 99.8%; Pred. No. 1.2e-20;
Matches 999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCCCTGTGGAGTGGAGGAGGCGCCGAGAACCTCAGAGATGGGACAGATACATCTCTG 60
Db 2256 GCCCTGTGGAGTGGAGGAGGCGCCGAGAACCTCAGAGATGGGACAGATACATCTCTG 2315
QY 61 AGGCAGGAGGGGACCAAGTGTGAGCTGTGAGCTGTGCGCTGGCCATGGCGAGCCCGGG 120
Db 2316 AGGCAGGAGGGGACCAAGTGTGAGCTGTGAGCTGTGCGCTGGCCATGGCGAGCCCGGG 2375
QY 121 GAGTACTGTGTGTGTGGGGCAGAGAGGACCTCAGCCACGCTCACCATCAGGGCTCTG 180
Db 2376 GAGTACTGTGTGTGTGGGGCAGAGAGGACCTCAGCCACGCTCACCATCAGGGCTCTG 2435
QY 181 CCTGCCAGTTTCATAGAAGATGTGAAAAACAGGAGGCCAGAGAGGGGCGCCACCGCTGTG 240
Db 2436 CCTGCCAGTTTCATAGAAGATGTGAAAAACAGGAGGCCAGAGAGGGGCGCCACCGCTGTG 2495
QY 241 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGGAGTGAGAGAGGGGTCTGAGACCTTC 300
Db 2496 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGGAGTGAGAGAGGGGTCTGAGACCTTC 2555

QY 301 AGAGATGGGACAGATACAGCCTTGAGCGAGGACGGACTAAATGTGAGCTGCAGATTCTGT 360
Db AGAGATGGGACAGATACAGCCTTGAGCGAGGACGGACTAAATGTGAGCTGCAGATTCTGT 2615
QY 361 GGCCTGGCCATGGCAGACACTGGGAGTACTGTGCTGTGCGGCGAGGAGGACCTCG 420
Db GGCCTGGCCATGGCAGACACTGGGAGTACTGTGCTGTGCGGCGAGGAGGACCTCG 2675
QY 421 GCTATGCTCACCGTCAAGGCTCTTACCCTCAAGTTTACAGAGGGTCTGAGGAACCAAGAG 480
Db GCTATGCTCACCGTCAAGGCTCTTACCCTCAAGTTTACAGAGGGTCTGAGGAACCAAGAG 2735
QY 481 GCCACAGAGGGGCAACACCGTGTGCGGTGTGAGCTGAGCAAGATGSCCCCGTGGAG 540
Db GCCACAGAGGGGCAACACCGTGTGCGGTGTGAGCTGAGCAAGATGSCCCCGTGGAG 2795
QY 541 TGTGTGAAGGGGCAAGACCGTGTGCGGTGTGAGCTGAGCAAGATGSCCCCGTGGAG 600
Db TGTGTGAAGGGGCAAGACCGTGTGCGGTGTGAGCTGAGCAAGATGSCCCCGTGGAG 2855
QY 601 GCCAGGTGTGAGCTGCAGATCCGCGGCTCTGTCGAGAGGACCGCTGGGAGTACCTGTGC 660
Db GCCAGGTGTGAGCTGCAGATCCGCGGCTCTGTCGAGAGGACCGCTGGGAGTACCTGTGC 2915
QY 661 ATGTGCGGGAAGCAGAGGACCTCAGCCTGTGAGCTGAGCAAGATGSCCCCGTGGAG 720
Db ATGTGCGGGAAGCAGAGGACCTCAGCCTGTGAGCTGAGCAAGATGSCCCCGTGGAG 2975
QY 721 ATAGAGGGTCTCAGGAATGAGAGGCCACAGAGGGGACCGCTGAGAGTGGGGAC 840
Db ATAGAGGGTCTCAGGAATGAGAGGCCACAGAGGGGACCGCTGAGAGTGGGGAC 780
QY 781 CTGACAAAGGGGACCGGTGAGTGGAGGAGGGGACCGCTCAGAGATGGGGAC 840
Db CTGACAAAGGGGACCGGTGAGTGGAGGAGGGGACCGCTCAGAGATGGGGAC 3035
QY 841 AGACACAGCCTGAGCAGGACGGTTCAGGTGTGAGCTGAGATCCGTGGCCTGGCTGTG 900
Db AGACACAGCCTGAGCAGGATGGGTTCAGGTGTGAGCTGAGATCCGTGGCCTGGCTGTG 3155
QY 901 GTGGATGCGGGGAGTACTCGTGTGTGTGCGGGCAGGAGGACCTCAGGCACACTCACT 960
Db GTGGATGCGGGGAGTACTCGTGTGTGTGCGGGCAGGAGGACCTCAGGCACACTCACT 3215
QY 961 GTCAGGCGCTGCTGCCAGATTATAGAAAGATGTGAAAA 1001
Db GTCAGGCGCTGCTGCCAGATTATAGAAAGATGTGAAAA 3256

RESULT 3
LOCUS CQ730656 18524 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 16590 from Patent WO02068579.
ACCESSION CQ730656
VERSION CQ730656.1 GI:42305092
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 16590 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
source 1..18524
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Query Match 71.3%; Score 714; DB 6; Length 18524;
Besc Local Similarity 83.5%; Pred. No. 3.1e-142;
Matches 810; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 1 GCCCTGTGGAGTGGAGAAAGGGGCCGAGAACTCTCAGAGATGGGACAGATACATCCTG 60
Db GCCCTGTGGAGTGGAGAAAGGGGCCGAGAACTCTCAGAGATGGGACAGATACATCCTG 11142
QY 61 AGGCGAGAGGGGACCAAGTGTGAGCTCAGATCTGTGCTGCGCTTGGCCATGGCGGACGCGGG 120
Db AGGCGAGAGGGGACCAAGTGTGAGCTCAGATCTGTGCTGCGCTTGGCCATGGCGGACGCGGG 11202
QY 121 GAGTACTTGTGTGTCGGGGCAGGAGGACCTCAGCAGCTCACCATCAGGGCTCTG 180
Db GAGTACTTGTGTGTCGGGGCAGGAGGACCTCAGCAGCTCACCATCAGGGCTCTG 11262
QY 181 CCTGCCAGGTTCATAGAAAGATGTGAAAAACAGAGAGGCCAGAGAGGGGCCACGGCTGTG 240
Db CCTGCCAGGTTCATAGAAAGATGTGAAAAACAGAGAGGCCAGAGAGGGGCCACGGCTGTG 11322
QY 241 CTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGTGAGTGGAGAAAGGGGTCTGAGACCTTC 300
Db CTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGTGAGTGGAGAAAGGGGTCTGAGACCTTC 11382
QY 301 AGAGATGGGACAGATACAGCCTGAGCAGGACGGGAGTGGCGAAGGGGTCTGAGACCTTC 11441
Db AGAGATGGGACAGATACAGCCTGAGCAGGACGGGAGTGGCGAAGGGGTCTGAGACCTTC 11442
QY 361 GGCTTGCCATGTGCAGACACTGGGGAGTACTCGTGTGTGTGCGGGCAGGAGAGCCTCG 420
Db GGCTTGCCATGTGCAGACACTGGGGAGTACTCGTGTGTGTGCGGGCAGGAGAGCCTCG 11502
QY 421 GCTATGCTCACCGTCAAGGCTCTTACCCTCAAGTTTGTGAGGGTCTGAGGAACGAGAG 480
Db GCTATGCTCACCGTCAAGGCTCTTACCCTCAAGTTTGTGAGGGTCTGAGGAACGAGAG 11562
QY 481 GCCACAGAAAGGGGCAACAGCCGTGCTGCGGTGTGAGCTGAGCAAGATGCCCCCGTGGAG 540
Db GCCACAGAAAGGGGCAACAGCCGTGCTGCGGTGTGAGCTGAGCAAGATGCCCCCGTGGAG 11622
QY 541 TGTGTGAAGGGGATGAGACCTCTCAGAGATGGAGACAGACAGCCTGAGGCGAGGACGGG 600
Db TGTGTGAAGGGGATGAGACCTCTCAGAGATGGAGACAGACAGCCTGAGGCGAGGACGGG 11682
QY 601 GCCAGGTGTGAGCTGCAGATCCGCGGCTCTGCGGAGAGGACGCTGGGAGTACCTGTGC 660
Db GCCAGGTGTGAGCTGCAGATCCGCGGCTCTGCGGAGAGGACGCTGGGAGTACCTGTGC 11742
QY 661 ATGTGCGGAAAGGAGAGGACCTCAGCCATGCTCACCGTCAGGGGCCATGCTTCCAGTTTC 720
Db ATGTGCGGAAAGGAGAGGACCTCAGCCATGCTCACCGTCAGGGGCCATGCTTCCAGTTTC 11802
QY 721 ATAGAGGTCTCAGGAATGAAGGGCCACAGAAAGGGGACACCGCCACGCTGTGTGTGAG 780
Db ATAGAGGTCTCAGGAATGAAGGGCCACAGAAAGGGGACACCGCCACGCTGTGTGTGAG 11862
QY 781 CTGAGCAAGCGCGGACCGGTGAGTGGAGAGGGGACATGAGACCTCAGAGATGGGGAC 840
Db CTGAGCAAGCGCGGACCGGTGAGTGGAGAGGGGACATGAGACCTCAGAGATGGGGAC 11922
QY 841 AGACACAGCCTGAGGAGGACCGGTTCAGGTGTGAGCTGAGATCCGTGGCCTGGCTGTG 900
Db AGACACAGCCTGAGGAGGACCGGTTCAGGTGTGAGCTGAGATCCGTGGCCTGGCTGTG 11982
QY 901 GTGGATGCGGGGAGTACTCGTGTGTGTGCGGGCAGGAGGACCTCAGCCACTCACT 960
Db GTGGATGCGGGGAGTACTCGTGTGTGTGCGGGCAGGAGGACCTCAGCCACTCACT 12042
QY 961 GTCAGGCGCC 970
Db GTCAGGCGCC 12111

```

RESULT 4
AX069754/c
LOCUS
DEFINITION Sequence 226 from Patent WO0102568.
ACCESSION AX069754
VERSION AX069754.1 GI:12579539
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klinger, J.,
Kasam, A., Reinhard, C., Randazzo, F., Kennedy, G.C., Pot, D.,
Lansom, G., Drmanac, R., Crkenjakov, R., Drmanac, S., Dickson, M.,
Labat, I., Leshkowitz, D., Kita, D., Garcia, V. and Strache-Crain, B.
HUMAN GENES AND GENE EXPRESSION PRODUCTS
Patent: WO 0102568-A 226 11-JAN-2001.
JOURNAL CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES
source
1. 352
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
misc_feature
1. 352
/note="n = A, T, C or G"
ORIGIN
Query Match 34.3%; Score 343.4; DB 6; Length 352;
Best Local Similarity 99.7%; Pred. No. 6.9e-63;
Matches 344; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 631 GTGGCAGAGGACGCTGGGAGTACTGCTGTCATGTGCGGGAAGGAGGACCTCAGCCATG 690
Db |
351 GTGGCAGAGGACGCTGGGAGTACTGCTGTCATGTGCGGGAAGGAGGACCTCAGCCATG 292
Oy 691 CTCACCGTCAGGCGCATGCTTCCAGTTTCATAGAGGTCTGAGGAATGAAGAGGCCACA 750
Db |
291 CTCACCGTCAGGCGCATGCTTCCAGTTTCATAGAGGTCTGAGGAATGAAGAGGCCACA 232
Oy 751 GAAGGGGACAGCGGCACGCTGCTGGTGTGAGCTGAGCAAGCGGCACCGTGGAGTGGAGG 810
Db |
231 GAAGGGGACAGCGGCACGCTGCTGGTGTGAGCTGAGCAAGCGGCACCGTGGAGTGGAGG 172
Oy 811 AAGGGGATGAGACCTTCAGAGATGGGACAGACACAGCCTCAGGACGAGCGGTCCAGG 870
Db |
171 AAGGGGATGAGACCTTCAGAGATGGGACAGACACAGCCTCAGGACGAGTGGTCCAGG 112
Oy 871 TGTGAGCTGAGATCCGTGGCTGCTGCTGAGTGCATGCGGGGAGTACTCTGTGTGTGTGC 930
Db |
111 TGTGAGCTGAGATCCGTGGCTGCTGCTGAGTGCATGCGGGGAGTACTCTGTGTGTGTGC 52
Oy 931 GGGCAGGAGAGGACCTCAGCCACACTCACTCTCAGGGCCCTGCCT 975
Db |
51 GGGCAGGAGAGGACCTCAGCCACACTCACTCTCAGGGCCCTGCCT 7

RESULT 5
HSA314905
LOCUS
DEFINITION Homo sapiens partial OBSCN gene for obscurin, exons 31-56.
ACCESSION AJ314905
VERSION AJ314905.1 GI:21104337
KEYWORDS OBSCN gene; obscurin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. Young, P., Ehler, B. and Gautel, M.
Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere
assembly

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 34667)
AUTHORS Gautel, M.S.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie,
Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Straasse
11, Dortmund, 44227, GERMANY
FEATURES
source
1. 34667
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
247..34433
/gene="OBSCN"
join(247..519,1332..1598,1817..2082,4388..4664,5087..5353,6008..6274,6375..6638,6939..7202,10212..10475,10837..11100,11667..11930,12466..12729,13094..13357,16768..17031,17573..17836,18196..18459,21989..22252,22617..22880,23476..23700,24686..24952,25215..25481,25573..25846,26423..26697,27413..27609,27798..27873,34161..34433)
/gene="OBSCN"
join(<247..519,1332..1598,1817..2082,4388..4664,5087..5353,6008..6274,6375..6638,6939..7202,10212..10475,10837..11100,11667..11930,12466..12729,13094..13357,16768..17031,17573..17836,18196..18459,21989..22252,22617..22880,23476..23700,24686..24952,25215..25481,25573..25846,26423..26697,27413..27609,27798..27873,34161..34433)
/codon_start=3
/product="obscurin"
/protein_id="CAC85752.1"
/db_xref="GI:21104338"
/db_xref="UniProt/Tx:EMBL:OBHNM3"
/translation="LHVGIITRLKMTVEVLGESCSFECVLSHESASDPAMTVGKTV
GSSRFQATQGRKYILVRENAPSADAGEVVSURGLTSKASLIVRERPAIILPLED
QWYPTVEASKSTASLHVREKANCFTTELNLQVEEKGTAVFTCKTBHPAIVTWKRG
LLELDVADGKQPSQEGTLRLTISALEKADSDTYTCDIQASQAQLLVQARRVHII
LELDVADGQSSATFCRISPNYEPVHWFDKTPLHANELNIDIAQPGVHVLTLR
PLVCKDSGTYFFAAGDORASAAALRVTEKPSVRELTADTITEGEDTLVCTSTCDI
QWCTKQDKTLRGSARCOLSHEGHAQLLTGATLQDSGRYKVCAGACSSSIVRHA
RPVRQALDKDLEVEGATLRCVLSVAAPVKWCYGNVNLVRGDKYSLRQEGMLE
LVRNLRPQSGRYSCSFQDQTSATLTVTALPAQFICKLNKATEGATATLRCELS
KAAPVWRKGSSETLRDGRYCLRDQGMALCELIQRLGLVADAAEYSCVGEATTSALT
IRPMPAFHIGLRHQSIEGATLRCELSKAAPVWRKGSRESLRDGRHSLRDQGA
CELIQCLGLVADAGEYSCVGEATTSALTIVKALPAKFTGLRNEEAVEGATMLWCE
LSKVAPEWRKGPENLRDGRYILRQEGTRCELIQCLGLVADAGEYLCVCGQERTSAT
LTIRALPARFIEDVKNQAREGATAVLQCELSNAAPEWRKGSSETLRDGRYSLRQDG
TKCELIQRLGLVADAGEYSCVCGQERTSATLTVRALPIKFTGLRNEEAEGATAVLR
CELSKMAPEWRKGPENLRDGRHSLRDQGMALCELIQRLGLVADAGEYLCVCGKERTS
AMLTVRAMPKSFIEGLRNEEATGDTATLWCELSKAAPEWRKGPENLRDGRHSLRQ
DGRCELIQRLGLVADAGEYSCVCGQERTSATLTVRALPARFIEDVKNQAREGATAV
LQCELSKAAPEWRKGSSETLRDGRYSLRQDGRTRCELIQHLGLVADAGEYSCVCGQER
TSATLTVRALPARFTQDKTREASEGATATLQCELSKVAPEWRKGPENLRDGRYSL
KQDGRTRCELIQHLGLVADAGEYSCVCGQERTSATLTVRALPARFTGLRNEEAEGAT
ATLQCELSKAAPEWRKGPENLRDGRYSLRQDGRTRCELIQHLGLVADAGEYLCVCGQ
SLKQDGTSCLELIQRLGLVADAGEYSCVCGQERTSATLTVRALPARFIEDVKNHATEG
ATAVLQCELSKAAPEWRKGSSETLRDGRYSLRQDGRTRCELIQRLGLVADAGEYLCV
QERTSATLTVRALPARFIEDVKNHATEGATATLHCELSKVAPEWRKGPENLRDGRS
RHSLRQDGRTRCELIQRLGLVADAGEYSCVCGQERTSATLTVRATEGATMLQCELSK
VAPVWRKGPENLRDGRYSLRQDGRTRCELIQHLGLVADAGEYSCVCGQERTSATLTV
KAPQVPEPLQSLQAEESTATLQCELSBPTATVMSKGLQVQAGREPRQCGCT
AEVLQDLOREDTGETCTCGSQATSLTVTAAPVRLRQHLQHEVDEGCTAHLCCCE
LSRAGSVWRKKSLLQIPPCAKYQWQDGAALLVRGVEQEDGATCDYCTDGTQSM
LSRVVRPRPKTRLRQLEQTDIALCCQLSDAESGAVQVWLKGEVGLRAGPKYEM
RSOGATRELLIHQLEAKDTGETACVTCGQKTAASLRVTGEPEVTIRGLVADAEVAD
DVFESCEVSRAGATGVQWCLQPLQNEVTEVAVRDRGIRHLKLVKTPPDAGTAFD
HLGNHASSAOLTVRAPVITILEPLQDVLQSEGODASFCQRLSRAGSGQARVWLGVP
QANEMNDITVEQGTLLHLLKVTLEDAGTVSFHVGTCSSSAQLKATKNTVRLGLE
VEALEGGEALFECLQSQPEVNAHTWLLDDPEVHTSENAEVFFENGLRHLILLNLRLP

```

```
QDSRVTFFLAGDMVTSFAFLTVR"
exon      247. .519
/gene="OBSCN"
number=31
intron    520. .1331
/gene="OBSCN"
number=31
exon      1332. .1598
/gene="OBSCN"
number=32
intron    1599. .1816
/gene="OBSCN"
number=32
exon      1817. .2082
/gene="OBSCN"
number=33
intron    2083. .4387
/gene="OBSCN"
number=33
exon      4388. .4664
/gene="OBSCN"
number=34
intron    4665. .5086
/gene="OBSCN"
number=34
exon      5087. .5353
/gene="OBSCN"
number=35
intron    5354. .6007
/gene="OBSCN"
number=35
exon      6008. .6274
/gene="OBSCN"
number=36
intron    6275. .6374
/gene="OBSCN"
number=36
exon      6375. .6638
/gene="OBSCN"
number=37
intron    6639. .6938
/gene="OBSCN"
number=37
exon      6939. .7202
/gene="OBSCN"
number=38
intron    7203. .10211
/gene="OBSCN"
number=38
exon      10212. .10475
/gene="OBSCN"
number=39
intron    10476. .10836
/gene="OBSCN"
number=39
exon      10837. .11100
/gene="OBSCN"
number=40
intron    11101. .11666
/gene="OBSCN"
number=40
exon      11667. .11930
/gene="OBSCN"
number=41
intron    11931. .12465
/gene="OBSCN"
number=41
exon      12466. .12729
/gene="OBSCN"
number=42
intron    12730. .13093
/gene="OBSCN"
number=42

13094. .13357
/gene="OBSCN"
number=43
intron    13358. .16767
/gene="OBSCN"
number=43
exon      16768. .17031
/gene="OBSCN"
number=44
intron    17032. .17572
/gene="OBSCN"
number=44
exon      17573. .17836
/gene="OBSCN"
number=45
intron    17837. .18195
/gene="OBSCN"
number=45
exon      18196. .18459
/gene="OBSCN"
number=46
intron    18460. .21988
/gene="OBSCN"
number=46
exon      21989. .22252
/gene="OBSCN"
number=47
intron    22253. .22616
/gene="OBSCN"
number=47
exon      22617. .22880
/gene="OBSCN"
number=48
intron    22881. .23475
/gene="OBSCN"
number=48
exon      23476. .23700
/gene="OBSCN"
number=49
intron    23701. .24685
/gene="OBSCN"

Query Match      33.5%; Score 335.2; DB 9; Length 34667;
Best Local Similarity 66.1%; Pred. No. 1.7e-61;
Matches 549; Conservative 0; Mismatches 243; Indels 38; Gaps 3;

Qy      175  GCTCTGCCTGCCAGGTTTCATAGAAGATGTGAAAAACACAGAGGCCAGAGAGGGGCCACG 234
Db      6374  GCCCTGCCCTGCCAGTTTCATCGGAAACTGAGAAAAACAGGAGGCCACAGAGGGGCCACG 6433

Qy      235  GCTGTGCTGCAGTGTGAGCTGAACAGTGCAGGCCCTGTGAGTGGAGTGGAGAAAGGGGTCTGAG 294
Db      6434  GCCACGCTCGGTGTGAGCTGAGCAAGGACGCCCTGTGAGTGGAGTGGAGAAAGGGGTCCGAG 6493

Qy      295  ACCCTCAGAGATGGGGACAGATACAGCCTGAGGCGAGGACGGGACTAAATGTGAGCTGCAG 354
Db      6494  ACCCTCAGAGATGGGGACAGATACAGCCTGAGGCGAGGACGGGCGCCATGTGTGAGCTGCAG 6553

Qy      355  ATTCTGGCCTGGCCATGCCACACTGGGGAGTACTCGTGTGTCGCGGGCAGAGAGG 414
Db      6554  ATCCGTGGCCTGGCCATGGTGGATCCCGGAGTACTCGTGTGTGTGTGAGAGAGAGG 6613

Qy      415  ACCTGGCTATGCTCACCGTCAGGCTCTACCCATCAAGTTCCACAGAGGGTCTGAGGAAC 474
Db      6614  ACCTCAGCCTCACTCACCATCAGGCGTAAGACCGTGTATCCAGAGCCGTGTCCGTGTCC 6673

Qy      475  GA-----AGAGGCCACAGAAAGGGGCAACAGCC 501
Db      6674  AATTTTTTACCTCTGCTGTGTCAGTCGTACCCATTCCTAGTCTACCTCTGCTGTGCCATGCC 6733

Qy      502  GTGCTCGGTGTGAGCTGAGCAAGATGCCCGCGGAGTGTGGAAGGGGATGAGACC 561
Db      6734  TTCCAGACACATCAGGGAATCCCTTCAGACTTTCTGGGCTGTGGAAGGAACCGAGGAAC 6793
```


Quality coverage: 5.55 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unkown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

*	1	2478:	contig	of 2478	bp in length
*	2479	2578:	gap	of unknown	length
*	2579	6948:	contig	of 4370	bp in length
*	6949	7048:	gap	of unknown	length
*	7049	9373:	contig	of 2325	bp in length
*	9374	9473:	gap	of unknown	length
*	9474	13303:	contig	of 3830	bp in length
*	13304	13403:	gap	of unknown	length
*	13404	18270:	contig	of 4867	bp in length
*	18271	18370:	gap	of unknown	length
*	18371	25589:	contig	of 7219	bp in length
*	25590	25689:	gap	of unknown	length
*	25690	34815:	contig	of 9126	bp in length
*	34816	34915:	gap	of unknown	length
*	34916	45742:	contig	of 10827	bp in length
*	45743	45842:	gap	of unknown	length
*	45843	62987:	contig	of 17145	bp in length
*	62988	63087:	gap	of unknown	length
*	63088	87584:	contig	of 24497	bp in length
*	87585	87684:	gap	of unknown	length
*	87685	114765:	contig	of 27081	bp in length
*	114766	114865:	gap	of unknown	length
*	114866	142334:	contig	of 27469	bp in length
*	142335	142434:	gap	of unknown	length
*	142435	174612:	contig	of 32178	bp in length

FEATURES

location/Qualifiers	1..174612	/organism="Homo sapiens"	
source		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
		/chromosome="1"	
		/clone="RP11-661B12"	
misc_feature	1..2478	/note="assembly_name:Contig4"	
misc_feature	2579..6948	/note="assembly_name:Contig5"	
misc_feature	7049..9373	/note="assembly_name:Contig6"	
misc_feature	9474..13303	/note="assembly_name:Contig7"	
misc_feature	13404..18270	/note="assembly_name:Contig8"	
misc_feature	18371..25589	/note="assembly_name:Contig9"	
misc_feature	25690..34815	/note="assembly_name:Contig10"	
misc_feature	34916..45742	/note="assembly_name:Contig11"	
misc_feature	45843..62987	/note="assembly_name:Contig12"	
	clone_end:SP6	vector_side:right"	
misc_feature	63088..87584	/note="assembly_name:Contig13"	
misc_feature	87685..114765	/note="assembly_name:Contig14"	
misc_feature	114866..142334	/note="assembly_name:Contig15"	
misc_feature	142435..174612	/note="assembly_name:Contig16"	

Query Match

32.1%; Score 321.6; DB 2; Length 174612;

Best Local Similarity 65.9%; Pred. No. 1e-58;		Matches 547; Conservative 0; Mismatches 244; Indels 39; Gaps 4;	
QY	175	GCTCTGCTCTGCCAGGTTTCATAGAAGATGTGAAAAACACGAGGCGCCACAGAGAAGGGCCACG	234
Db	174102	GCCCTGCTCTGCCAGTTTCATCGGAACTGAGNAACAGGAGCCACAGAGGGG-CACG	174044
QY	235	GCTGTGCTGCACTGTGAGCTGAACTGAGCGCCCTGTGGAATGGAGAAAGGGTCTGAG	294
Db	174043	GCCACGCTGCGGTGTGAGCTGAGCAAGACAGGCCCTGTGAGATGGAGAAAGGGTCCGAG	173984
QY	295	ACCCTCAGAGATGGGGACAGATACAGCCTTCAGCGAGGACGGACCTAAATGTTCAGCTGCAG	354
Db	173983	ACCCTCAGAGATGGGGACAGATACTGTCTGAGGCAAGCGGGCCATGTGTGAGCTGCAG	173924
QY	355	ATTCTGGCTGGCCATGTCAGACACACTGGGAGATCTCTGTGCTGTGCGGGCAGGAGAGG	414
Db	173923	ATCCGTGGCTGGCCATGGTGGATCCGCGGAGTACTCTGTGTGTGTGTGGAGAGGAGG	173864
QY	415	ACCTCGGTATGCTCACCGTCAGGGCTCTACCCATCAAGTTCCACAGAGGGTCTGAGGAAC	474
Db	173863	ACCTCAGCCTCACTCACCATCAGGGCGTAAGACCGTGTATCCAGAGCCGTGTCCGGTGTCC	173804
QY	475	GA-----AGAGGCCACAGAGGGGCAACAGCC	501
Db	173803	AATTTTTPACCTCTGCTGTCTGAGTCGTAGCCCATTTCTAGTCTACCTCTGAGTGTGCCATGCC	173744
QY	502	GTGTCGCGGTGTGAGCTCAGCAAGATGGCCCCCGTGGAGTGTGTGAAGGGGCATCAGACC	561
Db	173743	TTCCAGCACATCAGGHAATCCCTTCAGCTTCTTGGGCTGTGGAAGNAACAGAGGAAC	173684
QY	562	CTCAGAGATGAGACAGACACAGCCTTGAGGACGCGGGCCAGGTGTGAGCTGCAGATC	621
Db	173683	CGGGATTTGACGCGAGACTCCCACTTCAGGACACCTGGGACCTGTCTGTGGGAGTC	173624
QY	622	CGGCGCTCTGTGGCAGAGGACGTGGGGAGTACCTGTGCATGTGCGGCAAGGAGGACC	681
Db	173623	TGCTTATCCT-CGCTCTTGCAATGGCTGTCCCTGTGTTCCAGAACTGATCTCC	173565
QY	682	TCAGCCA----TGCTACCGTCTGAGGCGCATGCCTTTCCAAGTTTCATAGAGGCTGTAGAA	737
Db	173564	ATGTCTGTCTGTCTCATCTCTCCAGCGCATGCTGCCACCTTCATAGGAAGACTGAGACA	173505
QY	738	TGAAGAGGCCACAGAAGGGGACACGGCACCGCTGTGGTGTGAGCTGAGCAAGGGCGGACC	797
Db	173504	CCAAGAGCATAGAAAGGGCCAAGCCACGCTCGGTGTGAGCTGAGCAAGCGGCCCC	173445
QY	798	GCTGAGTGGAGGAAGGGCCATGAGACCTCAGAGATGGGACAGACACAGCCTGAGGCA	857
Db	173444	CGTGGATGGAGGAAGGGCGGTGAGAGCCTCAGAGATGGGACAGACATAGCCTGAGGCA	173385
QY	858	GGACGGGTTCAGGTGTGAGCTGACAGATCCGTGGCTGGCTGTGTGTGGATGCCGGGAGTA	917
Db	173384	GGACGGGGCTGTGTGCAAGCTGCAGATCTGTGGCTGGCTGTGCAAGATGCTGGGGAGTA	173325
QY	918	CTCGTGTGTGTGCGGGCAGGAGGACCTCAGCCACACTCACTGTCAAGG	967
Db	173324	CTCGTGTGTGTGGGAGGAGAGGACCTCTGCCACTCTCACCCTGTGAGG	173275

RESULT 9	AC026657/c
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

REFERENCE

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Waterston, R.H.
Unpublished
2 (bases 1 to 164766)
Waterston, R.H.
Direct Submission
Submitted (22-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7637349.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0245P10
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 139884 bases at least Q40
Consensus quality: 147686 bases at least Q30
Consensus quality: 151469 bases at least Q20
Insert size: 169000; agarose-fp
Insert size: 161074; sum-of-contigs
Quality coverage: 3.60 in Q20 bases; agarose-fp
Quality coverage: 3.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 1124: contig of 1124 bp in length
* 1125 1224: gap of unknown length
* 1225 3032: contig of 1808 bp in length
* 3033 3132: gap of unknown length
* 3133 4493: contig of 1361 bp in length
* 4494 4594: gap of unknown length
* 4594 5861: contig of 1267 bp in length
* 5861 5961: gap of unknown length
* 5961 7671: contig of 1710 bp in length
* 7671 9669: gap of unknown length
* 9669 9770: gap of unknown length
* 9770 11715: contig of 1946 bp in length
* 11716 11815: gap of unknown length
* 11816 14243: contig of 2428 bp in length
* 14244 14343: gap of unknown length
* 14344 16887: contig of 2544 bp in length
* 16888 16987: gap of unknown length
* 16988 19247: contig of 2260 bp in length
* 19248 19347: gap of unknown length
* 19348 21375: contig of 2028 bp in length
* 21376 21475: gap of unknown length
* 21476 25025: contig of 3550 bp in length
* 25026 25126: gap of unknown length
* 25126 28051: contig of 2926 bp in length
* 28052 28151: gap of unknown length
* 28152 32054: contig of 3903 bp in length
* 32055 32154: gap of unknown length
* 32155 35716: contig of 3562 bp in length
* 35717 35816: gap of unknown length
* 35817 40891: contig of 5075 bp in length
* 40892 40991: gap of unknown length
* 40992 44027: contig of 3036 bp in length
* 44028 44127: gap of unknown length

* 44128 48899: contig of 4772 bp in length
* 48900 48999: gap of unknown length
* 49000 54255: contig of 5256 bp in length
* 54256 54356: gap of unknown length
* 54357 58994: contig of 4639 bp in length
* 58995 59094: gap of unknown length
* 59095 64390: contig of 5296 bp in length
* 64391 64490: gap of unknown length
* 64491 70865: contig of 6375 bp in length
* 70866 70965: gap of unknown length
* 70966 78667: contig of 7702 bp in length
* 78668 78767: gap of unknown length
* 78768 87207: contig of 8440 bp in length
* 87208 87308: gap of unknown length
* 87309 97958: contig of 10551 bp in length
* 97959 109520: contig of 11562 bp in length
* 109521 109620: gap of unknown length
* 109621 120908: contig of 11288 bp in length
* 120909 121008: gap of unknown length
* 121009 141477: contig of 20469 bp in length
* 141478 141577: gap of unknown length
* 141578 162672: contig of 21095 bp in length
* 162673 162772: gap of unknown length
* 162773 163974: contig of 1202 bp in length
* 163975 164074: gap of unknown length
* 164075 164766: contig of 692 bp in length.

FEATURES
source
1. 164766
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-245P10"

1. 1124
/note="assembly_name:Contig14"
1225. 3032
/note="assembly_name:Contig21"
3133. 4493
/note="assembly_name:Contig23"
4594. 5860
/note="assembly_name:Contig24"
5961. 7670
/note="assembly_name:Contig25"
7771. 9669
/note="assembly_name:Contig26"
9770. 11715
/note="assembly_name:Contig27"
11816. 14243
/note="assembly_name:Contig28"
14344. 16887
/note="assembly_name:Contig29"
16988. 19247
/note="assembly_name:Contig30"
19348. 21375
/note="assembly_name:Contig31"
21476. 25025
/note="assembly_name:Contig32"
25126. 28051
/note="assembly_name:Contig33"
28152. 32054
/note="assembly_name:Contig34"
32155. 35716
/note="assembly_name:Contig35"
35817. 40891
/note="assembly_name:Contig36"
40992. 44027
/note="assembly_name:Contig37"
44128. 48899
/note="assembly_name:Contig38"
49000. 54255
/note="assembly_name:Contig39"
54356. 58994

misc_feature	/note="assembly_name:Contig40" 59095.. .64390 /note="assembly_name:Contig41" 64491.. .70865 /note="assembly_name:Contig42" 70966.. .78667 /note="assembly_name:Contig43" 78768.. .87207 /note="assembly_name:Contig44" 87308.. .97858 /note="assembly_name:Contig45" 97959.. .109520 /note="assembly_name:Contig46" 109621.. .120908 /note="assembly_name:Contig47" 121009.. .141477 /note="assembly_name:Contig48" 141578.. .162672 /note="assembly_name:Contig49" 162773.. .163974 /note="assembly_name:Contig9" 164075.. .164766 /note="assembly_name:Contig7"			
ORIGIN				
Query Match 27.8%; Score 278.6; DB 2; Length 164766; Best Local Similarity 62.6%; Pred. No. 1.5e-49; Matches 559; Conservative 0; Mismatches 234; Indels 100; Gaps 4;				
QY	175	GCTCTCCCTGCCAGGTT	CATAGACATGTGAAACACGAGGAGGCCACAGAGGGGCCACG	234
Db	133606	GCCTCCGCGCAGATT	CATAGAGATGTGAGAAATCATCGAGGCCACAGAGGGGCCACA	133547
QY	235	GCTGTGCTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGGAGTGGAGAAAGGGTCTGAG	294	
Db	133546	GCTGTGCTGCAGTGTGAGCTGAGCAAGGGGCCCGCTGGAGTGGCGAGAGGGTCTGAG	133487	
QY	295	ACCTCAGAGATGGGACAGATACAGCTGAGGCGAGGACGGGACTAAATGTGAGCTGCAG	354	
Db	133486	ACCCTCAGAGATGGGACAGATATAGCTGAGGCGAGGACGGGACAGGTGTGAGCTGCAG	133427	
QY	355	ATTGTCGGCTGGCCATGGCAGACACTGGGAGTACTCGTCGTGTCGGGCGAGGAGG	414	
Db	133426	ATTGTCGGCTGGCTGTGGAGGACACTGGAGATTTGTGTGTGTCGGGCGAGGAGAGA	133367	
QY	415	ACCTCGGCTATGCTCACCGTCAGGCTCTACCCATCAAGTTTCACAGAGGGTCTGAGGAAC	474	
Db	133366	ACCTCAGCTACACTCACTGTGAGGGTAAAGATCCTGTGTGGTCACATGACTTGTGGCT	133307	
QY	475	GAAGAGGCCACAGAGGGGGACA	-----GCCGTGCTGCGGTGT	513
Db	133306	TGGAATCTAAGTGTCTGTGTCTACCACCTTCTACTTCCAAATGTGGCACCCCTGTGGGA	133247	
QY	514	GAGCTGAGCAGATGGCCCGTGGAGTGTGGAGGGGCAT	-----GAGACCCCTCAG	566
Db	133246	ACCTTGAGGTGTGTGTCTTAAGTGGGTCTTCTGGGAGTCTCTGCGCTCTGCTTCTC	133187	
QY	567	AGATGGAGACACACAGCCTGAGGCGAGGACGGGGCCAGGTGTGAGCTGCAGATCCGCGG	626	
Db	133186	CTCATCAGGAGATGTCCTCTGTTTCATCCCATCCCTGCGTGTCTCTGTTCTCTCTTGGG	133127	
QY	627	CCTCGTGGCAGGACGCTGGGA	-----	650
Db	133126	CCATGTGCTTCTCGCTTCGTCATATTCATTGCTTATTTCTGCCATGGTCTCTTGGTGG	133067	
QY	651	-----GTACCTGTGCATGTGCGGAGAGGAGGACCTCAGCCATGCTCA	- 694	
Db	133066	TTTGTGTGGCTCTCTGCTGTGTCTGTCTGTCTCTCCACAGATTGTCCACAGATTCTGAT	133007	
QY	695	-----CCGTCAGGGCCATGCCCTTCCAAAGTTCATAGAGGGTCTGAG	734	
Db	133006	GATCTCAGTGACTGCTTGGTCTTCCAGCCCTGCCAGCCAGATTCTATAGACAAATGAC	132947	

Qy	735	GAATGAAGAGCCACAGAGGGGACACGGCCACGCTGTGTGTGCTGAGTGCACAGGGCGG 794
Db	132946	AAACAGAGAGCCACAGAGAGGGCCACGGCCACACTGCCTGTGAATGAGAGAGGTGGC 132887
Qy	795	ACCGTGTGAGTGGAGGAGGGGCATGAGACCTCAGAGATGGGACACAGACACACCTTGAG 854
Db	132886	CCCTGTGGAGTGGAGGAGGGACCTGAACCCCTCCGAGATGGGACACAGACACCTTGAG 132827
Qy	855	GCAGACGGTTCAGGTGTGAGCTGCAGATCCGTGGCTCGCTGTGTGTGATGCCGGGA 914
Db	132826	GCAGGATGGACACAGGTGTGAGCTGCAGATTCGTGGCTCTGTGTGGCAGATGCCGGGA 132767
Qy	915	GTACTCTGTGTGTGCGGCGAGGAGGACCTCAGGCACACTCACTCTCAGGG 967
Db	132766	GTACTCTGTGTGTGCGGCGAGGAGGACCTCAGGCACACTCAGCATCAGGG 132714
RESULT 10		
AC099089/c		
LOCUS		
Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.		
AC099089		
AC099089.5 GI:30522701		
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.		
Rattus norvegicus (Norway rat)		
Rattus norvegicus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
Rattus.		
1 (bases 1 to 260998)		
REFERENCE		
AUTHORS		
Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderson,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshuwa,L., Loundseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Pu,L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steelme,M., Strong,N., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,R., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,		

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 260998)
Worley, K. C.

Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 260998)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:22855456. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGLZ
Center clone name: CH230-154E3
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 216929 bases at least Q40
Consensus quality: 219228 bases at least Q30
Consensus quality: 220845 bases at least Q20
Estimated insert size: 226118; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 260998: contig of 260998 bp in length.
Location/Qualifiers
1..260998
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-154E3"
1..1486
/note="wgs_contig"

misc_feature
1..1486

FEATURES
source
1..260998
Location/Qualifiers
1..260998
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-154E3"
1..1486
/note="wgs_contig"

ORIGIN
Query Match 27.6%; Score 276; DB 2; Length 260998;
Best Local Similarity 61.1%; Pred. No. 5e-49;

Matches	485;	Conservative	0;	Mismatches	300;	Indels	9;	Gaps	2;
Qy	175	GCTCTGCTGCTCCAGGTTCTATAGAAAGATGTGTAAGAAACACGAGGAGCCAGAGAGGGCCACG		234					
Db	166307	GCCTCGTCTGCCAGTTCTATAGAAAGATGTGTAAGAAACACGAGGAGCCACAGAGGGACCATG		166248					
Qy	235	GCTGTGCTGCTGAGTGTGAGCTGAACAGTGCAGCCCTGTGGAGTGGAGAAAGGGGTCTGAG		294					
Db	166247	GCCACACTTCGGTGTGAGCTGACCAAGAGGCCCGGTGGATGGAGAAGGGAACAGAG		166188					
Qy	295	ACCTTCAGAGATGGGACAGATACAGCTCTGAGGAGGACGGGACTAAATGTGAGCTGCGAG		354					
Db	166187	ACCTTCAGAGATGGGACAGATACAGCTCTGAGGAGGACGGGACTAAATGTGAGCTGCGAG		166128					
Qy	355	ATTCGTGGCTTGGCCATGGCAGACACTCTGGGAGTACTCTGCTGCTGTGCGGCGCAGAGAGG		414					
Db	166127	ATCTGTAACCTGCTTGTGGCAGACGCGGGGAGTACTTGTGTGTATGCGGCGCAGAGAGG		166068					
Qy	415	ACCTCGGCTATGCTACCGTCTACCGTCTACCGTCTACCGTCTACCGTCTACCGTCTACCGT		473					
Db	166067	ACTTCAGCCACGCTGACTGTCAAGGGTAAGAGACACATGGCTTCCCGCCACCGCCATCC		166008					
Qy	474	CGAAGAGCCACAGAAAGGGGCAACAGCGCTCTCGGCTGTGAGCTGAGCAAGATGGCCCC		533					
Db	166007	CTGTGCACTTACTGTGTCATGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT		165949					
Qy	534	CGTGGAGTGTGGAAAGGGGATGAGACCTTCAGAGATGGAGACAGACAGACAGCCCTGAGCA		593					
Db	165948	-----AGAAAATACCAGCATCTGCAGCTGGCTTCCACCTCTCTGAGTGTGATGGGTTC		165896					
Qy	594	GGACGGGCGCAGGTGTGAGCTGAGATCCGGGCTCGTGGCAGAGGAGCGCTGGGGAGTA		653					
Db	165895	TTTCCTGTGTGCTGT		165836					
Qy	654	CCTGTGATGTGCGGGAAGAGAGAGCTCAGCCATGCTCAGCCGTGAGGGGCGATGCCCTTC		713					
Db	165835	CTTACCATGCTCTGTAGCAGATCTGATGTTCTTGTCTTCTTCTTCTTCTTCTTCTTCTT		165776					
Qy	714	CAAGTTCAAGAGGTCTGAGGAATGAGAGGCGCACAGAGGGGACACGGCCACGCTGTG		773					
Db	165775	CCACTTTATCAGAAAGGCTCAGGAAACAGAGGGGCGCACAGAGGGGACACGGTCACTGCA		165716					
Qy	774	GTGTGAGCTGAGCAAGGGGCGCACCGGTGTGAGTGGAGGAAGGGGCGATGAGCCCTCAGAGA		833					
Db	165715	GTGTGAACCTGAGCAAGGGGCGCACCGGTGTGAGTGGAGGAAGGGGCGATGAGCCCTCAGAGA		165656					
Qy	834	TGGGGACAGACACAGCTGAGGCGAGGCTCCAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTG		893					
Db	165655	TGGGGACAGACACAGCTGAGGCGAGGCTCCAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTG		165596					
Qy	894	GGCTGTGCTGATGCCGGGAGTACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT		953					
Db	165595	GGCTGTGAGCAGTCTGGAGAGTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT		165536					
Qy	954	ACTCACTGTGAGG		967					
Db	165535	ACTGACCATCAGG		165522					

RESULT 11
AL645854/c
LOCUS
DEFINITION Mouse DNA sequence from clone RP23-192P17 on chromosome 11,
complete sequence.
ACCESSION AL645854
VERSION AL645854.10 GI:20068636
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Smith, M.

Direct Submission
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:19699553.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-192P17 is from the RPCI-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6.

FEATURES

Location/Qualifiers
1..211829
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-192P17"
/clone_lib="RPCI-23"

ORIGIN

Query Match 26.7%; Score 267.2; DB 10; Length 211829;
Best Local Similarity 60.7%; Pred. No. 3.9e-47;
Matches 477; Conservative 0; Mismatches 298; Indels 11; Gaps 2;
QY 182 CTGCAGGTTTCATAGACATGTGAACACAGGAGGCGGACAGAGAGGCGGCTGTGC 241
Db 51183 CTGCCAGTTCGTAGGAACTGAGAAACAGGAGGCGGACAGAGAGGCGGCTGCAC 51124
QY 242 TGCAGTGTGAGCTGAACAGTGCAGCCCTGTGGAGTGGAGAAAGGGTCTTGAGACCCCTCA 301
Db 51123 TTCCGTGTGAGCTGACCAAGAGGCGGCGTGGAGTGGAGAGGAAGAACAGAGACCTGA 51064
QY 302 GAGATGGGGACAGATACAGCTGTAGGCGAGGACGGGACTAATGTGAGCTGCAGATTCGTG 361
Db 51063 GAAATGGGGACAAATACAGTCTGAAGCAGGATGGAGCTGTGTGTGAGCTGCAGATCTGTA 51004
QY 362 GCCTGGCCATGGCAGACATCGGGAGTACTCGTGGTGTGGCGGACGAGGAGGACCTCGG 421
Db 51003 GTCTGTGTGGCAGATCCAGGGAGTACTCGTGTGTGTGGCGGACGAGGAGGACTTCAG 50944
QY 422 CTATGCTCACCGTCAGGGCTTACCCATCAAGTTCACAGAGGGTCTGAGGAACGAAGAGG 481
Db 50943 CCAGCGTGACTGTCAAGGGT-----AAAGAGCGTGCATGGCTTCCTCCCGCCCTGCTGT 50888
QY 482 CCACAGAGGGGCAACAGCGTGTGCGGTGTGAGCTGAGCAGATGGCCCGCGTGGAGT 541
Db 50887 GTACTACTGTGCAATCCACCTTACTTCTTGTGAGTCATTATATATATGAACAGAAAAT 50828
QY 542 GGTGGAAGGGGCATGAGACCCCTCAGAGATGGAGACAGACACAGCGCTGAGGACAGACGGGG 601
Db 50827 ACCAGCATCTGCAGCTGGCTTCCCACTTTGAGTGTGTGGTGTCTCTCTCTTGTCTGC 50768
QY 602 CCAGGTGTGAGCTGCAGATCGGGGCGCTCGTGGCAGAGGACGCTGGGGAGTACTGTGCA 661
Db 50767 ATTCTATCTTGTTCCTTCATCC-----TGTTCATGTGCATTTCTATCTTCCCACTC 50715

QY 662 TGTGGGGAAGCAGAGAGGACCTCAGCCATGCTACCGTCAGGGCCATGCTTCCAAAGTTCA 721
Db 50714 TGTAGCAGATCTGATGCTCCAATCTTTGTCTTACCTCTAGCTCTGCTTGTCCACTTTG 50655
QY 722 TAGAGGGTCTCAGGAATGAAGAGGCCACAGAAAGGGACAGGCCACGCTGTGTGTGAGC 781
Db 50654 TCAGNAGGCTCAGAGGACNAGAGGCCACAGAAAGAGACACGACCACTGCAGTGTGAAC 50595
QY 782 TGACAAAGGGCGCACCGGTGGAGTGGAGAGGGGCATGAGACCCCTCAGAGATGGGACA 841
Db 50594 TGACAAAGGGCTGCCCGCTGGAGTGGAGGAAGGAAACAGAGACCTCAGAGATGGGACA 50535
QY 842 GACACAGCTGAGGACGAGGCGGTCCAGGTGTGAGCTGCAGATCCGTGGCTGTGGTGG 901
Db 50534 GATACAGCTGAAGCAGGATGGGGCTGTGTGTGAGCTGCAAAATCCCGAGGCTGACCATAG 50475
QY 902 TGGATGCGGGGAGTACTCGTGTGTGCGGGCAGGAGAGGACCTCAGCCACACTCACTG 961
Db 50474 CAGATGCTGGCGAGTACTTGTGCACGTTGTGACAGGAAAGACCTCAGCCACGCTGACCG 50415
QY 962 TCAGGG 967
Db 50414 TCAGGG 50409
RESULT 12
AX463538
LOCUS
DEFINITION Sequence 102 from Patent WO0248337.
ACCESSION AX463538
VERSION AX463538.1 GI:21886312
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
1
Duggan,B.M., Yao,M.G. and Griffin,J.A.
Secreted proteins
Patent: WO 0248337-A 102 20-JUN-2002;
INCYTE GENOMICS INC (US)
Location/Qualifiers
1..2155
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 2902793CB1"
ORIGIN
Query Match 23.9%; Score 238.8; DB 6; Length 2155;
Best Local Similarity 59.8%; Pred. No. 1e-40;
Matches 531; Conservative 0; Mismatches 262; Indels 95; Gaps 4;
QY 175 GCTCTGCTGCCAGGTTTCATAGAGATGTGAAAAACAGAGAGCCAGAGAGGCGCCACG 234
Db 205 GCGCTGCTGCCAGGTTTCACAGAGGTTCTGAGGAATGAAGAGGCCATGAAGGGGCCACA 264
QY 235 GCTGTGCTCAGTGTGAGCTGAACAGTGCAGCCCTGTGGAGTGGAGAAAGGGTCTGAG 294
Db 265 GCACACTGCAATGTAGCTGAGCAAGGAGCGCCCTGTGGAGTGGAGAAAGGGCCTTGAG 324
QY 295 ACCCTCAGAGATGGGGACAGATACAGCCCTGAGGCGAGGACGGGACTAAATGTGAGCTGAG 354
Db 325 GCTCTCAGAGATGGGGACAAATACAGCTGAGACAAGACGGGCTGTGTGTGAGCTGACG 384
QY 355 ATTCTGCGCTGGCCATGGCAGACACTGGGAGTACTCGTGTGTGCGGGCAGAGAGG 414
Db 385 ATTCTGCGCTGGCTATGGCAGATAACGGGGTGTACTCATGTGTGTGGGCGAGAGAGG 444
QY 415 ACCTCGGTATGCTCACCGTCAGGGCTCTACCCATCAAGTTTCACAGAGGGTC-----466
Db 445 ACCTCAGCTACACTCAGCTGTGAGGGGTAAAGATCCTATGTGGCCATGTGGGCTTGTGGCT 504

```
QY 467 TGAGAACGACGAGCCACAGAGGGGCAACAGCGTGTGCGGTGTGAGCTGAGCAAGA 526
Db 505 TGGGTATATACACCTCTCTGTGTGTCACCACTTCTGCCTCCAAATGGGCACATCTCTGTG 564
QY 527 TGGCCCCCGTGGAGTGGTGGGAAGGGGCATGAGACCTCTCAGAG- 568
Db 565 GAAACCTGTGATGTGTCTGTCTCTAACTGGGGCCCTCTAGGTATCTCTGCTCCCTT 624
QY 569 --ATGAGACACACACAGCCTGAGCAGGACGGGCCAGGTGTGAGCTGACAGATCCGGCG 626
Db 625 CTCATCAGTAGATGTCTCTCTCTCATGATGCTCTGTGTCTCTGTCTCTCTCTCTCTCT 684
QY 627 CCTCGTGCAGAGGCGCTGGGA----- 650
Db 685 CCAATTGGCTCTCTCATTTGTTATATATATTGCTCATCTAGCTATGGTCTCTTTGGTGTGT 744
QY 651 -----GTACCTGTGCATGTGCGGGAGGAGGAGGACCTTCAGCCATGCTCA----- 694
Db 745 GTGGCTCCTTATTGTGTCCATGTTCTGTCCGAAAATCCTCCACACAGTCTGATGATAT 804
QY 695 -----CCTCAGGCGCATGCTCTTCCAAATTCATAGAGGCTCTGAGGAATG 739
Db 805 CAGTGACTGTTTGGTCTCTCAGCGCTGCTGCGCAGATTTCATAGAGGATAGAGAAACC 864
QY 740 AAGAGCCACAGAGGGACACGGCCAGCTGTGTGTGAGCTGAGCAAGGGCGGACCGG 799
Db 865 AGAAGGCCACAGAGGGGCTCAGTCACATTCGAATGTAAGCTGAGAAAGCGGCGCCCG 924
QY 800 TGGAGTGAGGAAGGGGCATGAGACCTCAGAGATGGGACACACACAGCCTGAGGCGAG 859
Db 925 TGGAGTGAGAAAGGGGCCAACACCTCAAGATGGGACAGGTACAGCCTGAGGACAG 984
QY 860 ACGGCTCAGGTGTGAGCTGACAGATCCGTGGCTGGCTGTGTGTGAGTCCCGGGAGTACT 919
Db 985 ATGGACCAAGTGTGAGCTGACAGATTCGTGGCTGTGTGTGAGTCCCGGGAGTACT 1044
QY 920 CGTGTGTGCGGCAGAGAGGACCTCAGCCACTCAGCTGTGAGG 967
Db 1045 CGTGATATGTGAGCAGGAGGACCTCGGCCACGCTCAGTGTGAGG 1092
```

```
RESULT 13
HSM800639 HSM800639 2295 bp mRNA linear PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFZp434C153 (from clone DKFZp434C153);
DEFINITION partial cds.
ACCESSION AL080128
VERSION AL080128.1 GI:5262565
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 2295)
JOURNAL Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
COMMENT Direct Submission
Submitted (16-JUN-1999) MIPS, Am Klopferstritz 18a, D-82152
Martinsried, GERMANY
Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp434C153) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
```

FEATURES

```
source
1..2295
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp434C153"
```

```
/tissue_type="testis"
/clone_lib="434 (synonym: htes3) Vector pSport1; host
DH10B; sites NotI + Sali"
/dev stages="adult"
1..2295
/gene="DKFZp434C153"
<1..348
/gene="DKFZp434C153"
/notes="similarity to titin"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB45726_1"
/db_xref="GI:5262566"
/db_xref="UniProt/TREMBL:Q9Y4R5"
/translation="AMFCPPKPPDLSMISVTWVSFSAIPARPIEDMRNOKATEGATVT
LQCKLRKAAPVEMKRPNTLKDGRYSLKQDGTSCELQIRGLVIADAGEYSCICRQER
TSATLTVRGKDHM"
2232..2237
/polyA_signal /gene="DKFZp434C153"
2262
/polyA_site /gene="DKFZp434C153"
ORIGIN
```

```
Query Match 19.4%; Score 193.8; DB 9; Length 2295;
Best Local Similarity 78.8%; Pred. No. 4e-31; Mismatches 0; Gaps 0;
Matches 231; Conservative 0; Indels 0; Gaps 0;
QY 167 CCATCAGGCGCTCTGCTGCCAGGTTTCATAGAGATGTGAAAAACCCAGGAGCCAGAGAAG 226
Db 59 CTTCTCAGCCTCGCTGCCAGATTCATAGAGGATATGAGAAACCCAGGAGCCAGAGAAG 118
QY 227 GGGCCACGGCTGTGCTGCAGTGTGAGCTGAAACAGTGCAGCCCTGTGAGTGGAGAAAG 286
Db 119 GGGCTACAGTCACATTCGAATGTAAGCTGAGAAAGCGGCCCTGTGAGTGGAGAAAG 178
QY 287 GGTCTGAGACCTCAGAGATGGGACAGATACAGCTGAGCAGCAGGACCGGACTAAATGTG 346
Db 179 GGCCCAACACCTCAAGATGGGACAGGTACAGCTTGAAGCAGATGGGACCATGTGAGC 238
QY 347 AGCTGCAGATTCGTGGCCCTGGCCATGGCAGACACTGGGGAGTACTCTGCTGCTGCGGCG 406
Db 239 AGCTGCAGATTCGTGGCCCTGGCTCATAGCAGATCTGGAGATATCTCTGTCATATGTGAGC 298
QY 407 AGGAGAGACCTCGGCTATGCTACCGCTCAGGGCTTACCCATCAAGTTTCA 459
Db 299 AGGAGAGACCTCGGCCACGCTCACTGTGAGGGGTAAAGAGCCACATGTGACCA 351
```

RESULT 14

```
BV014682/c BV014682 621 bp DNA linear STS 30-MAY-2003
LOCUS S212P60282FH4.T0 C2ECHII/BI Mus musculus STS genomic, sequence
DEFINITION tagged site.
ACCESSION BV014682
VERSION BV014682.1 GI:31098577
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 621)
JOURNAL Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,
MEDLINE Lander,E.S., Lindblad-Toh,K. and Daly,M.J.
PUBMED The mosaic structure of variation in the laboratory mouse genome
12466852
COMMENT Nature 420 (6915), 574-578 (2002)
```

```
Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 16:50:48 ; Search time 3663.47 Seconds
(without alignments)
10400.615 Million cell updates/sec

Title: US-10-077-130-4_COPY_10500_11500

Perfect score: 1001
Sequence: 1 gcccttgagtgaggaa.....ttcatagaagatgtgaaaaa 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gse1:*

9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	377.4	37.7	595	2	BE086827 QV1-BT067
2	368.8	36.8	536	2	BF824937 ILO-HN003
3	340.6	34.0	486	4	BM030364 488988 MA
C 4	302	30.2	723	5	BQ446463 UI-H-EU1
C 5	278	27.8	487	2	BE466441 hz21h03.x
C 6	262.6	26.2	437	2	BF826423 CM4-HN002
7	218.2	21.8	369	2	BF826489 CM4-HN002
8	216	21.6	372	7	CV315157 CM1-HT087
C 9	196.2	19.6	570	9	CEB36029 tigr-gss-
C 10	194.4	19.4	359	2	BF853346 MR2-EN009
C 11	186	18.6	515	7	CV346932 MR2-EN038
C 12	175.2	17.5	601	2	BF398067 UI-R-B52-
C 13	173.2	17.3	324	5	BQ361458 PMO-OT023
C 14	173.2	17.3	690	8	BH025901 RPCI-24-2
C 15	147	14.7	204	4	BG987697 PM3-HT116
C 16	137.4	13.7	566	7	CN664157 A0809E09-
17	135.6	13.5	505	9	CE078765 tigr-gss-
18	135	13.5	1074	4	BJ730664 BJ730664
C 19	131	13.1	247	4	BG987696 PM3-HT116
C 20	129.8	13.0	742	7	CV352920 MR3-HT107
C 21	125.8	12.6	743	5	BV358283 603477905
C 22	125	12.5	647	8	AZ590254 1M0399B18
C 23	121.4	12.1	657	8	AZ506477 1M0347H09
24	115	11.5	628	8	AZ985514 2M0267G22

C 25	112.4	11.2	350	2	BF748073
C 26	108.4	10.8	380	4	BF957683
C 27	87.6	8.8	212	2	BF827521
C 28	81	8.1	382	1	AI478779
C 29	81	8.1	740	5	BU689712
C 30	78.6	7.9	330	8	AQ237886
C 31	72.8	7.3	510	5	BX478907
C 32	71.2	7.1	173	2	BE086762
C 33	69.2	6.9	1010	9	CNS02HPP
C 34	61.6	6.2	584	7	CK829379
C 35	60.4	6.0	899	9	CL073463
C 36	60.2	6.0	1970	9	CG748837
C 37	60.2	6.0	5558	3	AK029863
C 38	60	6.0	997	9	CNS05GPF5
C 39	58.4	5.8	895	2	BE731455
C 40	57.8	5.8	586	9	BX153687
C 41	57.8	5.8	1132	5	BQ401116
C 42	57.2	5.7	1153	8	CC252982
C 43	57	5.7	332	8	AZ754571
C 44	57	5.7	1393	9	CL043635
C 45	56.4	5.6	1184	5	BM911719

ALIGNMENTS

RESULT 1
BE086827
LOCUS QV1-BT0678-300400-182-f06 BT0678 Homo sapiens cDNA, mRNA sequence.
DEFINITION QV1-BT0678-300400-182-f06 BT0678 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE086827
VERSION BE086827.1 GI:8477221
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 595)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
E-mail: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=QV1-BT0678-300400-182-f06&t3=2000-04-30&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 595.
Location/Qualifiers
1. 595

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stages="Adult"
/clone_lib="BT0678"
/note="Organ: breast; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived

ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 1 (bases 1 to 486)
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
 Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C.,
 Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,
 Chitko-McKown, C.G., Perle, G., Holt, I., Karamecheva, S., Liang, P.,
 Quackenbush, J., and Keefe, J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013
 21182978
 CONTACT: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithem@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACAGTCAGCG
 Plate: 120 row: E column: 13
 Seq primer: ATTAGGTGCACTATAG.
 Location/Qualifiers
 1. 486
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 2B0V"
 /note="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."

FEATURES

source
 1. 486
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-EUI-azz-h-22-0-UI"
 /tissue_type="Osteoarthritic Cartilage"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Ctl"
 /note="Organ: Knee; Vector: p773-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 NCI CGAP Ctl is a normalized cDNA library containing the
 following tissue(s): Osteoarthritic Cartilage The library
 was constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dr primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into p773-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 TGATCAGCGT.
 TAG TISSUE=osteoarthritis cartilage
 TAG_LIB=UI-H-EUI
 TAG_SEQ=TGATCAGCGT"

ORIGIN

Query Match 34.0%; Score 340.6; DB 4; Length 486;
 Best Local Similarity 81.6%; Pred. No. 1e-68;
 Matches 394; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 Oy 515 AGCTGAGCAAGATGCCCCCGTGGAGTGTGTGAAGGGCATGAGACCTCAGAGATGGAG 574
 Db 1 AGCTGAGCAAGCGCCCCCGTGGAGTGTGTGAAGGGCATGAGACCTCAGAGATGGAG 60
 Oy 575 ACAGACACAGCCTGAGGCGAGGCGGGCGAGTGTGAGCTGCAGATCCGCGCCTCGTGG 634
 Db 61 ACAGACTCGGCTGAGCGAGGCGGGCGGTGTGTGAGTGGAAATCATGACCTGACCA 120
 Oy 635 CAGAGGACGCTGGGAGTACCTGTGCTATGTCGGGAAGAGAGACCTCAGCAGTCTCA 694
 Db 121 TGGAGGATGCTGGGAGTACTCATGATATGCGGCGAGGAGAAACCTCAGCAGCTCG 180
 Oy 695 CCGTCAGGCGCATGCTTCCAAAGTTTCATAGAGGCTCAGGATGAAGAGCCACAGAG 754
 Db 181 CCGTCAGGCGCCTGCTTCCAAAGTTTCATAGAGGCTCAGGATGAAGAGCCACAGAG 240
 Oy 755 GGGACACGGCCACGCTGTGTGAGCTGAGCAGGCGGCGCCCGGTGGAGTGGAGGAAGG 814
 Db 241 GGGCCACAGCCACTACTGCGCTGTGAGCTGAGCAGGCGGCGCCCGGTGGAGTGGAGGAAGG 300
 Oy 815 GGCATGAGACCTCAGAGATGGGACACACAGCCTCAGGCGAGCGGTCAGGTTGTG 874
 Db 301 GGGCCGAGACCTCAGAGCGGGGACAGAGTGGCGGTGAAGCAGGAGGAGCGGTGTG 360
 Oy 875 AGCTGCAATCCGTGGCTGCTGTGGTGGATCCGGGGAGTACTCGTGTGTGCGGGC 934

Db 361 AGCTGGAGATCGTGGGCTGACCGCTGGAGGATGCTGGGAGTACTCGTGCATGTGGGGC 420
 Oy 935 AGGAGAGACCTCAGCCACACTCACTGTGAGGGCCCTCGTCCAGATTCATAGAAGATG 994
 Db 421 AGGAGAGACCTCAGCCACAGTGGCGGTGAGGGCCCTCGTCCAGATTCCTGAAGGTC 480
 Oy 995 TCA 997
 Db 481 TCA 483

RESULT 4

LOCUS BQ446463/C
 DEFINITION UI-H-EUI-azz-h-22-0-UI.s1 NCI CGAP Ctl Homo sapiens cDNA clone
 UI-H-EUI-azz-h-22-0-UI 3', mRNA sequence.
 ACCESSION BQ446463
 VERSION BQ446463.1 GI:21249575
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 723)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 Seq primer: M13 FORWARD
 POLYA=yes.
 Location/Qualifiers
 1. 723
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-EUI-azz-h-22-0-UI"
 /tissue_type="Osteoarthritic Cartilage"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Ctl"
 /note="Organ: Knee; Vector: p773-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 NCI CGAP Ctl is a normalized cDNA library containing the
 following tissue(s): Osteoarthritic Cartilage The library
 was constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dr primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into p773-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 TGATCAGCGT.
 TAG TISSUE=osteoarthritis cartilage
 TAG_LIB=UI-H-EUI
 TAG_SEQ=TGATCAGCGT"

FEATURES

source
 1. 723
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-EUI-azz-h-22-0-UI"
 /tissue_type="Osteoarthritic Cartilage"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Ctl"
 /note="Organ: Knee; Vector: p773-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 NCI CGAP Ctl is a normalized cDNA library containing the
 following tissue(s): Osteoarthritic Cartilage The library
 was constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dr primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into p773-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 TGATCAGCGT.
 TAG TISSUE=osteoarthritis cartilage
 TAG_LIB=UI-H-EUI
 TAG_SEQ=TGATCAGCGT"

ORIGIN

Query Match 30.2%; Score 302; DB 5; Length 723;
 Best Local Similarity 68.0%; Pred. No. 1.1e-59;
 Matches 449; Conservative 0; Mismatches 207; Indels 4; Gaps 2;
 Oy 310 GACAGATACGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 369

Db 694 GCCCGGTGCAGCTGAGCCATGAGGGCCACCCGGGCCACAGCTGCTCATCTGCGGCCACCC 635
Qy 370 ATGCAGACACTGGGAGTACTCGTGTGTGGGCGAGAGAGGACCTCGGTATGCTC 429
Db 634 NTCAGGACAGTGGACGCTACAAGTGTAG-GCTGGGGCGCTGCAGCAGCTCCATGTC 576
Qy 430 ACCGTGAGGCTCTACCCATCAAGTGTACAGAGGCTGTGAGGAACAGAGGCCACAGAA 489
Db 575 AGGCTGATGCGGGCGCAGTGCCTCCAGAGGGCCCTGAAGGACCTCGAGGTGTGGAG 516
Qy 490 GGGGCAACAGCGTGTGCGGTGTGAGCTGAGCAAGATGGCC---CCCTGAGTGTGG 546
Db 515 GGTGTGTGCTCCACACTGCGCTGTGTGTGCTCATCTGTGGCTGCGCCGCTGAAGTGTGC 456
Qy 547 AAGGGCATGAGACCTCAGAGATGAGAGACAGACAGCTCAGGAGGAGCGGGCCAGG 606
Db 455 TATGAAACAAACGCTCTGAGGCCAGGTGACAAATACAGCTACGCCAGAGAGGTGCCATG 396
Qy 607 TGTGAGCTGAGATCCGGGCTGTGTGGCAGAGGAGCTGTGGGAGTACCTGTGCAATGTC 666
Db 395 CTGAGCTGGTGTGCGGAACCTCGCGCGCAGGACAGCGGGCGGTACTCATGCTCCTTC 336
Qy 667 GGGAGGAGAGGACCTCAGCCATGCTCACCGTCAGGGCCATCCCTTCCAAGTTCATAG 726
Db 335 GGGGACCACTACTTCTGCGACCTCAGAGTACTGCTCCCTGCTGCGCAGTTCATCGGG 276
Qy 727 GGTCTGAGGAATGAAGAGCCCAAGAGGGGACAGAGGCGGACAGCTGTGTGAGCTGAGC 786
Db 275 AAATGAGAAACAAAGAGGCCACAGAAAGGGCCACAGGCGCCACGNTCGGTGTGAGCTGAGC 216
Qy 787 AAGGGGACCGGTGAGTGGAGGAGGGGATGAGCCCTCAGAGATGGGGACAGAC 846
Db 215 AAGGAGGCCCCGTGTGAGTGGAGAAAGGGGTCCGAGAGCCCTCAGAGATGGGGACAGATAC 156
Qy 847 AGCCTGAGGACGAGCGGTCCAGGTGTGAGCTGAGATCCGTGGCTGCTGTGTGTGGAT 906
Db 155 TGTCTGAGGACGAGCGGGCCATGTGTGAGCTGAGATCCGTGGCTGCGCATGTGTGAT 96
Qy 907 GCGGGGAGTACTCGTGTGTGTGCGGGCAGGAGGAGCCTCAGCCACTCACTGTGAGG 966
Db 95 GCGCGGAGTACTCGTGTGTGTGTGGAGGAGGAGGACCTCAGCCTCACTCACCATCAGG 36

RESULT 5
LOCUS BE466441/c
DEFINITION hz21h03.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:3208661 3', similar to TR:Q10466 Q10466 TITIN, HEART ISOFORM N2-B ;, mRNA sequence.
ACCESSION BE466441
VERSION BE466441.1 GI:9512303
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 487)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco

FEATURES
source High quality sequence stop: 453.
Location/Qualifiers
1..487
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:3208661"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP GC6"
/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 27.8%; Score 278; DB 2; Length 487;
Best Local Similarity 77.2%; Pred. No. 4e-54;
Matches 338; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 529 GCCCCCCGTGAGTGGTGAAGGGGCATGAGACCTCAGAGATGGAGACACACAGCCTG 588
Db 457 GCGCCCTGGAAGTGGTGTATGAACACAGCTCTGAGGCCAGTGACAATACAGCTA 398
Qy 589 AGGAGAGACGGGCGCAGGTGTGAGCTGCAGATCCGGGCTCTGGCAGAGGACGCTGGG 648
Db 397 CGCCAGGAGGTGCCATGCTGGAGCTGGTCCGGAACCTCCGCGCAGGACAGCGGG 338
Qy 649 GAGTACTGTGCATGTGCGGAGGAGAGGACCTCAGCCATGCTACCGTCAGGGCCATG 708
Db 337 CGGTACTCATGCTCTTCGGGGACAGACTACTTCTGCCACCTCAGTGACTGCGCCTG 278
Qy 709 CCTTCCAAAGTTTCATAGAGGTCTGAGGAATGAAGAGGCCACAGAGGGGACACGGCCACG 768
Db 277 CCTGCCCAGTTTCATCGGGAACCTGAGAACACAGGAGGCCACAGAGGGGCCACGCACG 218
Qy 769 CTGTGTGTGAGCTGAGCAAGCGGCGCACCGGTGAGTGGAGGAAGGGGCATGAGACCTTC 828
Db 217 CTGCGGTGTGAGCTGAGCAAGCGGCGCAGCCCTGTGGAGTGGAGAAAGGGTCCGAGACCTC 158
Qy 829 AGAGATGGGACAGACACAGCCTTGAGCGAGGAGCGGTCCAGGTGTGAGCTGAGATCCCT 888
Db 157 AGAGATGGGACAGATGCTGTCTGAGGACAGGAGCGGCCATGTGTGAGCTGAGATCCCT 98
Qy 889 GGCTCGCTGTGTGTGATCCGGGGAGTACTCGTGTGTGTGGCGGAGGAGGACCTCA 948
Db 97 GGCTGGCCATGGTGGATCCGCGGAGTACTCGTGTGTGTGTGGAGGAGGAGGACCTCA 38
Qy 949 GCCACACTCACTGTTCAGG 966
Db 37 GCGTCACTCACCATCAGG 20

RESULT 6
LOCUS BF826423/c
DEFINITION CM4-HN0020-221100-451-a02 HN0020 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF826423
VERSION BF826423.1 GI:12169626
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 437)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Brentani,R.R., Reis,L.F., de Souza,S.J., and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 10737800
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-HN0020-
 221100-451-a02&t3=2000-11-22&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 19
 High quality sequence stop: 437.

FEATURES

Location/Qualifiers
 1. 437
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="HN0020"
 /note="Organ: head normal; Vector: puc18; Site 1: Sma1;
 Site 2: Sma1; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

ORIGIN

Query Match 26.2%; Score 262.6; DB 2; Length 437;
 Best Local Similarity 80.4%; Pred. No. 1.5e-50;
 Matches 332; Conservative 0; Mismatches 79; Indels 2; Gaps 2;
 Oy 149 GGACCTGAGCCACGCTACCATCAGGCTCTGCTGCCAGGTTTCATAGAAATGTGAAA 208
 Db 425 GGTCTTACGCTACACTCACTGTGAGGGCCCTGCTGCCAGACTCATAGAGGATATGAGA 366
 Oy 209 ACCAGGAGCCAGAGAAGGGCCACGGCTGTGCTGAGTGTGAGTGAACAGTGCAGGCC 268
 Db 365 ACCAGAAGCCACAGAAGGGGCTCAGTCACATTGCAATGTAGCTGAGAAGGGCGCCC 306
 Oy 269 CTGTGGAGTGGAGAAAGGGGTCTGAGACCTTCAGAGATGGGACAGATACAGCCTGAGGC 328
 Db 305 CCGTGGAGTGGAGAAAGGGGCCAACCCCTCAAGATGGGACAGGTACAGCCTGAAGC 246
 Oy 329 AGGACGGACTAAATGTGAGCTGCAGATTCGTGGCTGGCCATGCGACACACTGGGGAGT 388
 Db 245 AGGATGGACCAAGTTGTGAGCTGCAGATTCGTGGCTGGTTCATAGCAGATCTGGAGAT 186
 Oy 389 ACTGTGCGTGTGGGGCAGGAGGACCTCGGCTATGCTCACCGTCAGGCTCTTACCCA 448
 Db 185 ACTGTGCATATGTAGCAGGAGAGGACCTCGGCCACGCCCACTGTTCAGGGC-CTGCCGG 127
 Oy 449 TCAAGTTTCACAGAGGTTCTGAGGAACGAAGAGGGCCACAGAAGGGGCCAACGCCGTGTGC 508
 Db 126 CCAGATTTCATAGAAGTGTGAGAAATCACGAGGCCACAGAAGGGGCCACACAGCTGTGTCG 67
 Oy 509 GGTGTGAGCTAGCAAGATGCCCCCGTGGAGTGTGGTGAAGGGGCATGAGACC 561
 Db 66 AGTGTGAGCTGAGCACG-CGCCCCCGTGGAGTGGCGAAGGGGTCTGAACC 15

RESULT 7

BF826489

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

1. 369

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="HN0020"

/note="Organ: head normal; Vector: puc18; Site 1: Sma1;
 Site 2: Sma1; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

Query Match 21.8%; Score 218.2; DB 2; Length 369;
 Best Local Similarity 81.4%; Pred. No. 3.3e-40;
 Matches 253; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
 Oy 2 CCCCTGTGGAGTGGAGGAAGGGGCCCGAGAACCTCAGAGATGGGACAGATACATCTGA 61
 Db 52 CCCCGTTGGAGTGGATGAAGGGGCCCAACACCTCAAAGATGGACAGGTACAGCCTGA 111
 Oy 62 GGCAGGAGGGACACAGGTGTGAGCTGCAGATCTGTGGCTGGCCATGGCGACCGCGGG 121
 Db 112 AGCAGGATGGGACCAAGTTGTGAGCTGCAGATTCGTGGCTGGTTCATAGCAGATGCTGAG 171
 Oy 122 AGTACTTGTGTGTGTCGGGCAGGAGGACCTCAGCCACGCTCACCACATCAGGCTCTGC 181
 Db 172 AATACTCGTCATATGTGAGCAGGAGAGGACCTCGGCCACGCTCACTGTCTGGGCCCTGC 231
 Oy 182 CTGCCAGGTTTCATAGAAGATGTGAAAAACAGGAGGGCCAGAGAAGGGGCCACGGTGTGC 241
 Db 232 CGGCCAGATTTCATAGATATGTGAGAAATCACGAGGCCACAGAGGGGCCACAGCTGTGC 291

```

QY 242 TGCAAGTGTGAGCTGAACAGTGCAGCCCTGTGTGAGTGGAGAAAGGGGTCTGAGACCTCA 301
Db 292 TGCAGTGTGAGCTGAGCAGAGCGGCCCTGTGTGAGTGGCGGAGGGGTCTGAGACCTCA 351
QY 302 GAGATGGGAC 312
Db 352 GAGCTGAGGAC 362

RESULT 8
CV315157 372 bp mRNA linear EST 24-SEP-2004
LOCUS CMI-HT0877-210900-418-b07 HT0877 Homo sapiens cDNA, mRNA sequence.
ACCESSION CV315157
VERSION CV315157.1 GI:52638371
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bal,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. Location/Qualifiers
FEATURES
source
1..372
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0877"
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 21.6%; Score 216; DB 7; Length 372;
Best Local Similarity 84.7%; Pred. No. 1.1e-39;
Matches 254; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

QY 703 GCCATGCTTCCAAAGTTTCATAG-AGGGTCTGAGGAATGAGAGGCCACAGAGGGGACAC 761
Db 65 GCCCTGTGGCCGAGTTTCATAGACGATGTGAGAAATCAGAGGCCACAGAGGGGCCAC 124
QY 762 GGCCACGCTGTGTGTGAGCTGAGCAAGGGCGGACCCGCTGAGTGGAGGAAGGGGATGA 821
Db 125 AGCTGTGCTGCAGTGTGAGCTGACCAAGGGCGGCCCTGTGAGTGGCGGAGGGTCTGA 184
QY 822 GACCTCAGAGATGGGACACAGACAGCCTGAGCAGGACGGGTCCAGGTGTGAGCTGCA 881
Db 185 GACCTCAGAGATGGGACACAGATATCGCTGAGCAGGACGGGACGAGGTGTGAGCTGCA 244

```

```

QY 882 GATCGTGGCCTGGCTGTGTGGATGCGGGGAGTACTCTGTGTGTGTCGGGCGAGGAGAG 941
Db 245 GATTGTGGCTGGCTGTGTGGAGGACACTGGAGAGTATTTGTGTGTGTCGGGCGAGGAG 304
QY 942 GACCTCAGCCACACTCACTGTTCAGGGCCCTGCCCTGCCAGATTTCATAGAAGATGTGAAAAA 1001
Db 305 AACCTCAGCTACACTCACTGTTCAGGGCCCTGCCAGCCAGATTTCATAGACACATGACAAA 364

RESULT 9
CV316029 570 bp DNA linear GSS 30-SEP-2003
LOCUS tigr-gss-dog-17000332435861 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE836029
VERSION CE836029.1 GI:37177049
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 570)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES
source
1..570
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"
ORIGIN
Query Match 19.6%; Score 196.2; DB 9; Length 570;
Best Local Similarity 83.8%; Pred. No. 4.7e-35;
Matches 222; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 703 GCATGCGCTTCCAAAGTTTCATAGAGGTCTGAGGAATGAAGAGCCACAGAGGGGACAG 762
Db 376 GGCTCTCGCCCAAGTTTCAACCAAGGCCCTGAGGAAGAGGAGGCCACGGAAGGGCCAG 317
QY 763 GCCACGCTGTGTGTGAGCTGAGCAAGCGCGCACCGGTGGAGTGGAGGAAGGGGATGAG 822
Db 316 GCCACGCTGTGTGTGAGCTGAGCAAGCGCGGCCCTGTGAGTGGAGGAAGGGGCTGAG 257
QY 823 ACCTCAGAGATGGGACACAGACAGCCTTGAAGGACGGGTCCAGGTGTGAGCTGACG 882
Db 256 GGCCTCAGAGCGGGGACAGAGGGTTCAGCTTGAAGGACGGGCGGCGCTGTGCGAGCTG 197
QY 883 ATCCGTGGCCTGGCTGTGTGTGATGCGGGGAGTACTCGTGTGTGCGGGCAGGAGG 942
Db 196 ATCCGGGGCCTGGCGTGTGAGGACTCCGGGGAGTACTCGTGTGTGCGGGCAGGAGG 137
QY 943 ACCTCAGCCACACTCACTGTTCAGGG 967
Db 136 ACGGGCGCTGTGCTCACCATCAGGG 112

```

RESULT 10
 BF853346/c
 LOCUS
 DEFINITION MR2-EN0091-191200-004-f08 EN0091 Homo sapiens CDNA, mRNA sequence. EST 16-JAN-2001
 ACCESSION BF853346
 VERSION BF853346.1 GI:12241090
 EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 359)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Negai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Mateukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 PUBMED 10737800
 COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR2&t2=MR2-EN0091-191200-004-f08&t3=2000-12-19&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 19
 High quality sequence stop: 358.
 Location/Qualifiers
 1. 359
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="EN0091"
 /note="Organ: lung normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES

source

1. 515
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="BN0387"
 /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 19.4%; Score 194.4; DB 2; Length 359;
 Best Local Similarity 82.5%; Pred. No. 1.2e-34;
 Matches 222; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 199 GATGTGAACACAGGAGCCAGAGAGGGCCACGGCTGTGCTGACGTGAGCTGAC 258
 DB 359 GATGTGAGAATACGAGGCCACAGAGAGGGGCCACAGCTGTGCTGACGTGAGCTGAGC 300
 QY 259 AGTCAGACCCCTGTGGAGTGAGAAAGGGGTCTCAGACCCCTCAGAGATGGGGACAGATAC 318
 DB 299 AAGNCGGCCCCGTGGAGTGGCGAGGGGTCTGAGACCCCTCAGAGATGGGGACAGATAT 240
 QY 319 AGCCTGAGGACGAGCGGACGACTTAATGTGAGCTGACAGATTCTGGCCTGGCCATGGCAGAC 378
 DB 239 AGCCTGAGGACGAGCGGACGAGGTGTGAGCTGCAGATTCTGGCCTGGCTGTGGAGGAC 180
 QY 379 ACTGGGAGTACTCGTGGCTGTGGGCGAGAGGACCTCGGTATGCTCACCGTCAGG 438

Db 179 ACTGGAGAGTATTTGTGTGTGCGGGCAGAGAGAACTCAGCTACACTCACTGTCTCAGG 120
 QY 439 GCTCTACCCATCAAGTTCACAGAGGGTCT 467
 Db 119 GGTAAAGATCCTGTGTGTGTCACATGGACT 91

RESULT 11

CV346932
 LOCUS
 DEFINITION MR2-EN0387-290900-001-g11 BN0387 Homo sapiens CDNA, mRNA sequence.
 ACCESSION CV346932
 VERSION CV346932.1 GI:52670146
 EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 515)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Negai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Mateukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 PUBMED 10737800
 COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. http://www.ludwig.org.br.
 Location/Qualifiers
 1. 515
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="BN0387"
 /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES
 source
 1. 515
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="BN0387"
 /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
 Query Match 18.6%; Score 186; DB 7; Length 515;
 Best Local Similarity 78.7%; Pred. No. 1.1e-32;
 Matches 222; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 686 CCATGCTCACCGTCAGGGCCATGCTTCCAAGTTTCATAGAGGGTCTGAGGATGAGAGG 745
 DB 172 CCATGTGACCCCTTCTAGCCCTGCTGCACGATTCATCAAGATCTGAAGACCAAGAGG 231
 QY 746 CCACAGAAGGGGACACGGCCACCGCTGTGTGAGCTGAGCAAGCGGACCGTGGAGT 805
 DB 232 CCTCAGAAGGGGCCACAGCTACACTGCAGTGTGAGCTGAGCAAGGTGGCCCTGTGGAT 291
 QY 806 GGAGGAAGGGGCAATGAGACCTTCAGAGATGGGGACAGACACAGCTGAGGACGAGGT 865
 DB 292 GGAAGAAGGGTCTCTGAGACCCCTCAGAGATGGGGCGAGATACAGCTCCTGAAGCAGGATGGA 351
 QY 866 CCAGGTGTGAGCTGCAGATCCGTGGCTGGTGTGGTGGATGCCGGGGAGTACTCTGTGTG 925

Db 352 CGAGCGCTGAGCTGCAGATCCATGACCTGTCCTGTGGCGGATGCTGGGAATACTCATGCA 411
 QY 926 TGTGCGGGCAGGAGAGACCTCAGCCACACTCACTGTCTCAGGG 967
 Db 412 TGTGTGACAAGAGAGAGACCTCGGCCACGCTCAGGTCTCAGGG 453

RESULT 12
 BF398067/c
 LOCUS
 DEFINITION
 UI-R-BS2-beg-h-06-0-UI.s1 UI-R-BS2 Rattus norvegicus cDNA clone
 UI-R-BS2-beg-h-06-0-UI 3', mRNA sequence.
 BF398067
 ACCESSION
 BF398067
 VERSION
 BF398067.1 GI:11383052
 EST.
 KEYWORDS
 Rattus norvegicus (Norway rat)

ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 601)

REFERENCE
 AUTHORS
 Title
 Ronaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: bento-soares@uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 normalized embryo at 13 dpc library cDNA Library preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES

Location/Qualifiers
 1..601
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BS2-beg-h-06-0-UI"
 /dev_stage="embryonic 13 dpc"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-BS2"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BS2
 library is a subtracted library derived from 13 dpc whole
 embryo tissue. For a detailed description of the library
 from which this clone was derived, please visit our web
 site at ratest.eng.uiowa.edu. The subtraction has been
 previously described in (Bonaldo, Lennon and Soares,
 Genome Research 6:791-806, 1996)
 TAG_1152=embryo at 13 dpc
 TAG_L1B=UI-R-BS2
 TAG_SEQ=AATCC"

ORIGIN

Query Match 17.5%; Score 175.2; DB 2; Length 601;
 Best Local Similarity 71.3%; Pred. No. 3.7e-30;
 Matches 231; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
 QY 644 CTGGGGAGTACCTGTGATGTGGCGGAAGAGAGACCTCAGCCATGCTCACCGTCAGGG 703

Db 517 CTAGGGCGCTCCATGTTCTCAGAGGGTGTGATTGTGTCTGTGTCTTTCTGACCTCCCCAG 458
 QY 704 CCATGCCCTTCCAAGTTTCATAGAGGTTCTGAGGAATGAAGAGGCCACAGAGGGGACACGG 763
 Db 457 CCCITCCGGCTAAGTTTCATAGAAAGTCTGAAGATGAAGGGGCCACAGAGGAACCAACAG 398
 QY 764 CCACGCTGTGTGTGAGCTGAGCAAGCGGCACCGGTGGAGTGGAGGAAGGGGCATGAGA 823
 Db 397 CCACGCTGAGCTGCAAACTGAGCAAGCGGCTCCGGTGAAGCTGGAAGAGAGGAACAAGA 338
 QY 824 CCCTCAGAGATGGGGCAGACACACAGCCCTGAGGAGGAGCGGTCCAGTGTGAGCTCAGA 883
 Db 337 CTTGCGAGACGGAGACAATATGCGCTGAGGACGAGCGAGCTGTGTGAGCTGCAGA 278
 QY 884 TCCGTGCGCTGTGTGTGTGGATGCCGGGGAGTACTCGTGTGTGTGCGGCGAGGAGAGA 943
 Db 277 TCCGTGCGCTGACACAGCCGATGCTGGGAGTACTCATGTGTGTGTGGCAGGAGAGA 218
 QY 944 CTTGAGCCACTCCTCCTGTCTCAGGG 967
 Db 217 CATCAGCTGCTGACTGTCAAGG 194

RESULT 13

BQ361458/c

LOCUS

DEFINITION

PMO-OT0232-240501-005-a12 OT0232 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BQ361458

VERSION

BQ361458.1 GI:21031179

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 324)

Nagai, M., E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,

O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PMO&t2=PMO-OT0232-

240501-005-a12&t3=2001-05-24&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 21

High quality sequence stop: 324.

Location/Qualifiers

1..324

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="OT0232"

/note="Organ: ovary; Vector: puc18; Site 1: SmaI; Site 2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No.

196,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

```

ORIGIN
Query Match          17.3%; Score 173.2; DB 5; Length 324;
Best Local Similarity 89.0%; Pred. No. 9.9e-30;
Matches 187; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 757 GACAGCGCCAGCTGTGTGAGTGAGTGAGCGGCAACGGTGGAGGAGGGG 816
Db 324 GCCAGCGCCAGCTGTGTGAGTGAGTGAGCGGCAACGGTGGAGGAGGGG 265

Qy 817 CATGAGACCTTCAGAGATGGGACACAGACAGCCCTGAGGAGGAGGAGGAG 876
Db 264 TCCGAGACCTTCAGAGATGGGACACAGATCTGTCTGAGGAGGAGGAGGAG 205

Qy 877 CTGAGATCCGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 936
Db 204 CTGAGATCCGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 145

Qy 937 GAGAGGACTCAGCCACACTACTCTCAGG 966
Db 144 GAGAGGACTCAGCCCTACTCACCATCAGG 115

```

```

RESULT 14
BH025901/c
LOCUS
DEFINITION
  RPCI-24-259H14-TV RPCI-24 Mus musculus genomic clone
  RPCI-24-259H14, genomic survey sequence.
ACCESSION
  BH025901
VERSION
  BH025901.1 GI:14789365
KEYWORDS
  GSS.
SOURCE
  Mus musculus (house mouse)
  Mus musculus
  ORGANISM
    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
    1 (bases 1 to 690)
    Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
    Tsagay, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
    Russell, D., de Jong, P. and Fraser, C.M.
    Mouse BAC End Sequences from Library RPCI-24
    Unpublished (1999)
    Other GSSs: RPCI-24-259H14.TJ
    Contact: Shaying Zhao
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850, USA
    Tel: 301 838 0200
    Fax: 301 838 0208
    Email: szhao@tigr.org
    Clones are derived from the mouse BAC library RPCI-24. For BAC
    library availability, please contact Pieter de Jong
    (pjejong@mail.cho.org). Clones may be purchased from BACPAC
    Resources (http://www.chori.org/bacpac/orderingframe.html). BAC end
    page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
    Plate: 259 row: H column: 14
    Seq primer: T7
    Class: BAC ends.

```

```

FEATURES
  source
    1..690
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="RPCI-24-259H14"
    /sex="Male"
    /cell_type="Spleen/Brain"
    /clone_lib="RPCI-24"
    /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
    RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
    library was cloned in the pTARBAC1 cloning vector at the

```

BamHI sites using MboI partially digested male C57BL/6J DNA."

```

ORIGIN
Query Match          17.3%; Score 173.2; DB 8; Length 690;
Best Local Similarity 82.7%; Pred. No. 1.1e-29;
Matches 210; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

Qy 703 GCCATGCTTCCAAAGTTTCATAGAGGGTCTGAGGAATGAAGAGCCACAGAGGGGACAG 762
Db 253 GCCCTGCCCTCCAGATTCATAGAGATTGAGAGTCAAGAGGCCACAGAGGCCACCATG 194

Qy 763 GCCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 822
Db 193 GCCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 134

Qy 823 ACCCTCAGAGATGGGACAGACACAGCCCTGAGGAGGAGCGGCTCCAGGTGTGAGCTGAC 882
Db 133 ACCCTGAGAGATGGGACAGATACAGCTGAGGAGGAGTGGCTGTGTGTGTGTGTGTGTGT 74

Qy 883 ATCGGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 942
Db 73 ATCTGTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 15

Qy 943 ACCTCAGCCACACT 956
Db 14 ACCTCAGCCACACT 1

```

```

RESULT 15
BG987697/c
LOCUS
DEFINITION
  PM3-HT1165-110101-001-e07 HT1165 Homo sapiens cDNA, mRNA sequence.
ACCESSION
  BG987697
VERSION
  BG987697.1 GI:14391767
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 204)
    Dias Neto, E., Garcia Corrae, R., Vertovski-Almeida, S., Briones, M.R.,
    Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Coeta, P.F.,
    Goldman, G.H., Carvalho, A.P., Matsukuma, A., Baia, G.S., Simpson, D.H.,
    Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
    O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
    Simpson, A.J.
    Shotgun sequencing of the human transcriptome with ORF expressed
    sequence tags

```

```

TITLE
  Shotgun sequencing of the human transcriptome with ORF expressed
  sequence tags
JOURNAL
  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE
  20202663
PUBMED
  10737800
COMMENT
  Contact: Simpson A.J.G.
  Laboratory of Cancer Genetics
  Ludwig Institute for Cancer Research
  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
  Brazil
  Tel: +55-11-2704922
  Fax: +55-11-2707001
  Email: asimpson@ludwig.org.br
  This sequence was derived from the FAPESP/LICR Human Cancer Genome
  Project. This entry can be seen in the following URL
  (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM3&t2=PM3-HT1165-
  110101-001-e07&t3=2001-01-11&t4=1)
  Seq primer: puc 18 forward
  High quality sequence start: 35
  High quality sequence stop: 204.
  Location/Qualifiers
    1..204
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /dev_stage="Adult"

```

/clone_lib="HT1165"
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match	14.7%	Score 147;	DB 4;	Length 204;
Best Local Similarity	82.8%	Pred. No. 1.2e-23;		
Matches 168;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;
Qy	57	CCTGAGGCGAGGAGGACACGAGTGTGAGCTGCAGATCTGTGGCTGGCCATGGCGGACGC	116	
Db	204	CCTGAGGCGAGGAGGAGGAGTGTGAGCTGCAGATCTGTGGCTGGCTGTGGAGGACAC	145	
Qy	117	CGGGGAGTACTTGTGTGTGTGTGGGCGAGGAGGACCTCAGCCACGCTCACCATCAGGGC	176	
Db	144	TGGAGAGTATTGTGTGTGTGTGGGCGAGGAGGACCTCAGCTACACTCACTTTTCAGGGC	85	
Qy	177	TCTGCCTGCCAGCTTCATAGAGATGTGAAACACGAGGAGGCGCAGAGAGGGGCCACGGC	236	
Db	84	CCTGCCAGCCAGATTTCATACCCCATGACCAACAGGAAGCCAGAGAGGGGCCACGGC	25	
Qy	237	TGTGCTGCAGTGTGAGCTGAACA	259	
Db	24	CACACTGCACTGTGAACCTGAGCA	2	

Search completed: March 21, 2005, 14:13:38
Job time : 3668.47 secs

Result No.	Score	Query Match	Length	DB	ID	Description	
1	1001	100.0	3225	6	AX542969	Sequence	
2	1001	100.0	4041	9	AB046859	Homo sapi	
3	1001	100.0	4175	6	AX039409	Sequence	
4	1001	100.0	4936	6	AX207401	Sequence	
5	1001	100.0	5454	6	AX504255	Sequence	
6	999.4	99.8	5007	6	AX039410	Sequence	
7	999.4	99.8	5207	6	AX255867	Sequence	
8	999.4	99.8	5207	6	AR442826	Sequence	
9	999.4	99.8	5207	6	AR442827	Sequence	
10	999.4	99.8	5207	6	AX430856	Sequence	
11	999.4	99.8	5207	6	AX430858	Sequence	
12	999.4	99.8	7928	6	AX039412	Sequence	
13	754.6	75.4	18524	6	CQ730656	Sequence	
14	738.4	73.8	4740	10	AY603754	Mus muscu	
15	406.6	40.6	1183	10	BC046431	Mus muscu	
16	243	24.3	62164	9	AL670729	Human DNA	
17	243	24.3	174612	2	AC023889	Homo sapi	
18	184.2	18.4	5923	5	AY603753	Danio rer	
19	181	18.1	164766	2	AC026657	Homo sapi	

Qy	481	TACAGCAGCCCTCGAGCAAAAGTCTCTCTGGAGGGCCAGCCACCTGGCTCTGAGGAG	540
Db	3058	TACAGCAGCCCTCGAGCAAAAGTCTCTCTGGAGGGCCAGCCACCTGGCTCTGAGGAG	3117
Qy	541	GAGAGCCAGGGGCGGTGAGCCAAACCCCTGCCCCAGCAAAAGACCTTTCGATTCCAGACA	600
Db	3118	GAGAGCCAGGGGCGGTGAGCCAAACCCCTGCCCCAGCAAAAGACCTTTCGATTCCAGACA	3177
Qy	601	CAGATCCAGAGGGGCGGCTTCAGCGTGGTGGCGCAATCTCTGGAGAAAGGCCAGGGGCGG	660
Db	3178	CAGATCCAGAGGGGCGGCTTCAGCGTGGTGGCGCAATCTCTGGAGAAAGGCCAGGGGCGG	3237
Qy	661	CGCTGGCGCGCAAGATCATCCCCCTACACCCAAAGACAAGACAGCAGTGTGCGCGAA	720
Db	3238	CGCTGGCGCGCAAGATCATCCCCCTACACCCAAAGACAAGACAGCAGTGTGCGCGAA	3297
Qy	721	TACAGAGCCCTCAAGGGCGTGGCCACCCGCACTCTGGCCCCAGCTGCACGAGCCTACCTC	780
Db	3298	TACAGAGCCCTCAAGGGCGTGGCCACCCGCACTCTGGCCCCAGCTGCACGAGCCTACCTC	3357
Qy	781	AGCCCCGGCACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCCGAGCTGTCCCGCTGC	840
Db	3358	AGCCCCGGCACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCCGAGCTGTCCCGCTGC	3417
Qy	841	CTGGCCGAGAGGGGCTCCTTACTCAGAATCCGAGTGAAGACTACTCTGTGGCAGATGTTG	900
Db	3418	CTGGCCGAGAGGGGCTCCTTACTCAGAATCCGAGTGAAGACTACTCTGTGGCAGATGTTG	3477
Qy	901	AGTGCCACCCAGTACTGCAACACAGCACATCCTGCACTGGACCTTGAGGTCGGAGAAC	960
Db	3478	AGTGCCACCCAGTACTGCAACACAGCACATCCTGCACTGGACCTTGAGGTCGGAGAAC	3537
Qy	961	ATGATCATACCGAATACACCTGCTCAAGTCGTGGAACCT	1001
Db	3538	ATGATCATACCGAATACACCTGCTCAAGTCGTGGAACCT	3578

```

RESULT 4
AX207401
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
    source
        1. .4936
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

```

QY	121	AACCTCCAGCTTCTGACCATCCTGGTGTTGGTGTGCTGAGCACTGGGTGTGTATACACTGC	180
Df	3499	AACCTTCAGCTTCTGACCATCCTGGTGTTGGTGTGCTGAGCACTGGGTGTGTATACACTGC	3558
QY	181	AGCGTAGCAAAATGGCGCTGGGACAGTACCACACACGGGCCCTCTCCGGAAAGGCAGAGCGC	240
Df	3559	AGCGTAGCAAAATGGCGCTGGGACAGTACCACACACGGGCCCTCTCCGGAAAGGCAGAGCGC	3618
QY	241	CCCTCATCTTCGCCATGCCCCGGATATCGGGAGGTTGTAACGCGAATGGGTGCTGTGGTTC	300
Df	3619	CCCTCATCTTCGCCATGCCCCGGATATCGGGAGGTTGTAACGCGAATGGGTGCTGTGGTTC	3678
QY	301	TGGAAAGCCCGTGAATCTACGGGCCCTGTGACCTACATTGTGTGAGTGCAGCTAGAAAGC	360
Df	3679	TGGAAAGCCCGTGAATCTACGGGCCCTGTGACCTACATTGTGTGAGTGCAGCTAGAAAGC	3738
QY	361	GGCAGCTGGACCACTCTGGGCTCCGACATCTTTGACTGTGCTTACCTGACCAAGCAAGTTC	420
Df	3739	GGCAGCTGGACCACTCTGGGCTCCGACATCTTTGACTGTGCTTACCTGACCAAGCAAGTTC	3798
QY	421	TCCCAGGGGTGGCACCTACACTTCCGACGGCATGTGTGAGCAAGGCAGGAATGGGTCCC	480
Df	3799	TCCCAGGGGTGGCACCTACACTTCCGACGGCATGTGTGAGCAAGGCAGGAATGGGTCCC	3858
QY	481	TACAGCAGCCCCCTCGGAGCAAGTCTCTCTGGAGGGCCCCAGCCACTGCGCTCTGAGAG	540
Df	3859	TACAGCAGCCCCCTCGGAGCAAGTCTCTCTGGAGGGCCCCAGCCACTGCGCTCTGAGAG	3918
QY	541	GAGAGCAGAGGGGGGTGAGCCCAACCCCTGCCAGACAAGAAGCTTTGGCAATTCAGACA	600
Df	3919	GAGAGCAGAGGGGGGTGAGCCCAACCCCTGCCAGACAAGAAGCTTTGGCAATTCAGACA	3978
QY	601	CAGATCCAGAGGGGGCGCTTTCAGCGTGTGCGGCAATGCTGGGAGAACGCCAGCGGGCGG	660
Df	3979	CAGATCCAGAGGGGGCGCTTTCAGCGTGTGCGGCAATGCTGGGAGAACGCCAGCGGGCGG	4038
QY	661	CGCTGCGGCCCAAGATCATCCCTTACCACCCCAAGGACAAGCAGCAGTGTCTGGCGCAA	720
Df	4039	CGCTGCGGCCCAAGATCATCCCTTACCACCCCAAGGACAAGCAGCAGTGTCTGGCGCAA	4098
QY	721	TAGAGGCCCTCAAGGGCGTGGCCACCGCAGCTTGGCCAGCTGCAAGCAGCTTACCTC	780
Df	4099	TAGAGGCCCTCAAGGGCGTGGCCACCGCAGCTTGGCCAGCTGCAAGCAGCTTACCTC	4158
QY	781	AGCCCCCGGCACCTGGTGTCTATCTTGGAGCTGTGCTCTGGGCCCGAGCTGTCCCTCTC	840
Df	4159	AGCCCCCGGCACCTGGTGTCTATCTTGGAGCTGTGCTCTGGGCCCGAGCTGTCCCTCTC	4218
QY	841	CTGGCCGAGAGGGCCCTTACTCAGAAATCCGAGGTGAAGGACTACTGTGGCAGATGTTG	900
Df	4219	CTGGCCGAGAGGGCCCTTACTCAGAAATCCGAGGTGAAGGACTACTGTGGCAGATGTTG	4278
QY	901	AGTGCCACCCAGTACTCTGCACAAACAGCAGCATCTTGCACCTTGGACCTCAGGTCGAGAAC	960
Df	4279	AGTGCCACCCAGTACTCTGCACAAACAGCAGCATCTTGCACCTTGGACCTCAGGTCGAGAAC	4338
QY	961	ATGATCATCACCGAATAACAACCTGTCTCAAGGTCTGGACCT	1001
Df	4339	ATGATCATCACCGAATAACAACCTGTCTCAAGGTCTGGACCT	4379
RESULT 5			
AX504255			
LOCUS	AX504255	5454 bp	DNA linear PAT 27-SEP-2002
DEFINITION	Sequence 44 from Patent WO0233099.		
ACCESSION	AX504255		
KEYWORDS	AX504255.1 GI:23386095		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		

AUTHORS	Gururajan, R., Baughn, M.R., Walia, N.K., Elliott, V.S., Xu, Y., Arvizu, C., Yao, M.G., Ramkumar, J., Ding, L., Tang, Y.T., Hafalia, A.J., Nguyen, D.B., Gandhi, A.R., Lu, Y., Yue, H., Burford, N., Bandman, O., Tribouley, C.M., Lai, P.G., Recipon, S.A., Lu, D.A., Borowsky, M.L., Thornton, M., Swarnaker, A., Thangavelu, K., Khan, F.A. and Ison, C.H.											
TITLE	Human kinases											
JOURNAL	Patent: WO 0233099-A 44 25-APR-2002;											
FEATURES	Incyte Genomics, Inc. (US)											
source	Location/Qualifiers											
	1. 5454											
	/organism="Homo sapiens"											
	/mol_type="unassigned DNA"											
	/db_xref="taxon:9606"											
	/notes="Incyte ID No: 7638121CB1"											
ORIGIN												
	Query Match 100.0%; Score 1001; DB 6; Length 5454;											
	Best Local Similarity 100.0%; Pred. No. 4.7e-156;											
	Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;											
Qy	1	GGCCAGTCA	GTGACAC	TGGCGT	CGCCAG	TGTG	CAGCCAG	CGCAG	GTGTC	CGCCAG	CGACCTCGG	60
Db	3892	GGCCAGTCA	GTGACAC	TGGCGT	CGCCAG	TGTG	CAGCCAG	CGCAG	GTGTC	CGCCAG	CGACCTCGG	3951
Qy	61	AGCAAGAC	GGAGG	CCCCCT	TGGAG	AGCAG	CGCGT	GTCT	CTAT	CTCTG	CGCACCTCAAG	120
Db	3952	AGCAAGAC	GGAGG	CCCCCT	TGGAG	AGCAG	CGCGT	GTCT	CTAT	CTCTG	CGCACCTCAAG	4011
Qy	121	AACCTTCAG	CTTCTG	ACCATC	CTGTTGG	TGGTGG	CTGAG	CACCTGG	GTGTG	TACACCTGC	180	
Db	4012	AACCTTCAG	CTTCTG	ACCATC	CTGTTGG	TGGTGG	CTGAG	CACCTGG	GTGTG	TACACCTGC	4071	
Qy	181	AGCGTAGCA	ATGCG	CTGGG	AGCAG	TGAC	CACAC	CGGCG	CTCT	C	CGGAAGGCAGAGCGC	240
Db	4072	AGCGTAGCA	ATGCG	CTGGG	AGCAG	TGAC	CACAC	CGGCG	CTCT	C	CGGAAGGCAGAGCGC	4131
Qy	241	CCCTCAT	CTTTG	CGCATC	CCCGG	ATAT	CGGGAG	AGGTG	TAC	CGGAGTGGGTG	CTGCTGGTC	300
Db	4132	CCCTCAT	CTTTG	CGCATC	CCCGG	ATAT	CGGGAG	AGGTG	TAC	CGGAGTGGGTG	CTGCTGGTC	4191
Qy	301	TGGNAGCC	CGTGG	NAATC	CTTAGG	CCCTG	TGAC	CTTAC	ATTG	TGCAGTGC	AGCCTAGAGGC	360
Db	4192	TGGNAGCC	CGTGG	NAATC	CTTAGG	CCCTG	TGAC	CTTAC	ATTG	TGCAGTGC	AGCCTAGAGGC	4251
Qy	361	GGCAGCT	TGGAC	CACTG	GGCTC	CGACAT	CTTTG	GACTG	CTGT	ACCTG	ACCGACGACGCTC	420
Db	4252	GGCAGCT	TGGAC	CACTG	GGCTC	CGACAT	CTTTG	GACTG	CTGT	ACCTG	ACCGACGACGCTC	4311
Qy	421	TCCCGGGT	TGGCAC	CTTAC	CTTCC	GAC	GGCAT	TGCT	TCAG	CAAGG	CAGGAAATGGGTCCC	480
Db	4312	TCCCGGGT	TGGCAC	CTTAC	CTTCC	GAC	GGCAT	TGCT	TCAG	CAAGG	CAGGAAATGGGTCCC	4371
Qy	481	TACAGCAG	CCCCCT	CGGAG	CAAGT	CTCT	CTGG	AGGGCC	ACG	CAC	CTGGGCTCTGAGGAG	540
Db	4372	TACAGCAG	CCCCCT	CGGAG	CAAGT	CTCT	CTGG	AGGGCC	ACG	CAC	CTGGGCTCTGAGGAG	4431
Qy	541	GAGAGC	CAGGG	CGGT	CAG	CCCA	ACCCCT	TGCC	CAG	CAAA	AGACCTTCG	600
Db	4432	GAGAGC	CAGGG	CGGT	CAG	CCCA	ACCCCT	TGCC	CAG	CAAA	AGACCTTCG	4491
Qy	601	CAGATCC	AGAGGG	CGCTTT	CAG	CGTGGT	TGCG	CAATG	CTGG	GAGAA	AGGCCAGCGGCGG	660
Db	4492	CAGATCC	AGAGGG	CGCTTT	CAG	CGTGGT	TGCG	CAATG	CTGG	GAGAA	AGGCCAGCGGCGG	4551
Qy	661	CGCGTGC	CGGCA	GATCAT	CCCC	CTAC	CCCC	AGGAC	AGAC	CAG	CAGTGTCTGCGGCA	720
Db	4552	CGCGTGC	CGGCA	GATCAT	CCCC	CTAC	CCCC	AGGAC	AGAC	CAG	CAGTGTCTGCGGCA	4611
Qy	721	TACGAG	CCCTCA	AGGG	CGCTGG	CCAC	CCG	CACTGG	CC	CACTG	CAAGGCTTACCTC	780
Db	4612	TACGAG	CCCTCA	AGGG	CGCTGG	CCAC	CCG	CACTGG	CC	CACTG	CAAGGCTTACCTC	4671
Qy	781	AGCCCC	CGGAC	CTGG	TGCTCAT	CTTGG	AGCTGTG	CTCT	TGG	GGCCG	CCGAGCTGCTCCCCCTGC	840

Query Match 99.8%; Score 999.4; DB 6; Length 5007;
Best Local Similarity 99.9%; Pred. No. 8.8e-156;
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCAGTCAGTGACACTGGCTCCAGAGTGTACGCCAGCAGCTGCCAGGCCACCTGG 60
Db |||||
Qy 3451 GGCAGTCAGTGACACTGGCTCCAGAGTGTACGCCAGCAGCTGCCAGGCCACCTGG 3510
Db |||||
Qy 61 AGCAAGAGCGAGGCCCTCCAGAGAGCAGCAGCCGTGCTCATCTCTGCCACCTCAAG 120
Db |||||
Qy 3511 AGCAAGAGCGAGGCCCTCCAGAGAGCAGCAGCCGTGCTCATCTCTGCCACCTCAAG 3570
Db |||||
Qy 121 AACTTCAGCTTCTGACCAATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
Db |||||
Qy 3571 AACTTCAGCTTCTGACCAATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3630
Db |||||
Qy 181 AGCGTCAGCAATGCGCTGGGAGCAGTGACCAACCGGGGCTCTCCGGAAGCAGAGCG 240
Db |||||
Qy 3631 AGCGTCAGCAATGCGCTGGGAGCAGTGACCAACCGGGGCTCTCCGGAAGCAGAGCG 3690
Db |||||
Qy 241 CCTCATCTTCGCGCATGCCCGGATATCGGGAGGTGTACCGGATGGGTGCTGTGTC 300
Db |||||
Qy 3691 CCTCATCTTCGCGCATGCCCGGATATCGGGAGGTGTACCGGATGGGTGCTGTGTC 3750
Db |||||
Qy 301 TGGAGCCCGTGGAAATCTTACCGGCCCTGTGACCTATTGTGCACTAGAGGC 360
Db |||||
Qy 3751 TGGAGCCCGTGGAAATCTTACCGGCCCTGTGACCTATTGTGCACTAGAGGC 3810
Db |||||
Qy 361 GCGAGCTGACACACACTGCGCTCCGACATCTTTGACTGCTGCTGCTGCTGCTGCTG 420
Db |||||
Qy 3811 GCGAGCTGACACACACTGCGCTCCGACATCTTTGACTGCTGCTGCTGCTGCTGCTG 3870
Db |||||
Qy 421 TCCCGGGTGGCACTTACACCTTCCGCGAGGATGTGTGAGCAAGGAGGATGGGTCCC 480
Db |||||
Qy 3871 TCCCGGGTGGCACTTACACCTTCCGCGAGGATGTGTGAGCAAGGAGGATGGGTCCC 3930
Db |||||
Qy 481 TACAGAGCCCTCGGAGCAAGTCTCTCGGAGGGCCAGCCACTTGGCTCTGAGGAG 540
Db |||||
Qy 3931 TACAGAGCCCTCGGAGCAAGTCTCTCGGAGGGCCAGCCACTTGGCTCTGAGGAG 3990
Db |||||
Qy 541 GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTCGCAATCCAGACA 600
Db |||||
Qy 3991 GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTCGCAATCCAGACA 4050
Db |||||
Qy 601 GAGATCCAGAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTCGCAATCCAGACA 660
Db |||||
Qy 4051 GAGATCCAGAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTCGCAATCCAGACA 4110
Db |||||
Qy 661 GCGTGGCGCCCAAGATCATCCCTTACCAACCCAGGACAGACAGTGTCTGCGGAA 720
Db |||||
Qy 4111 GCGTGGCGCCCAAGATCATCCCTTACCAACCCAGGACAGACAGTGTCTGCGGAA 4170
Db |||||
Qy 721 TACAGAGCCCTCAAGGGCTCGGCCACCCGACCTGGGCCAGCTGCAAGCAGCTACTCTC 780
Db |||||
Qy 4171 TACAGAGCCCTCAAGGGCTCGGCCACCCGACCTGGGCCAGCTGCAAGCAGCTACTCTC 4230
Db |||||
Qy 781 AGCCCCCGGCACTGGTGTCTCATCTTGAGTGTGCTCTGGGCCGAGCTGCTCCCTGC 840
Db |||||
Qy 4231 AGCCCCCGGCACTGGTGTCTCATCTTGAGTGTGCTCTGGGCCGAGCTGCTCCCTGC 4290
Db |||||
Qy 841 CTGGCCGAGAGGGCTCTCTACTCAGATCCAGAGTGGAGTACTACTGTGGCAGATTTG 900
Db |||||
Qy 4291 CTGGCCGAGAGGGCTCTCTACTCAGATCCAGAGTGGAGTACTACTGTGGCAGATTTG 4350
Db |||||
Qy 901 AGTGCCACCCAGTACTGTCACAAACAGCACAATCTGCACTGGACCTGAGTCTCGGAAC 960
Db |||||
Qy 4351 AGTGCCACCCAGTACTGTCACAAACAGCACAATCTGCACTGGACCTGAGTCTCGGAAC 4410
Db |||||
Qy 961 ATGATCATCAGCAATACAACTGCTCAAGGTCTGGAACCT 1001
Db |||||
Qy 4411 ATGATCATCAGCAATACAACTGCTCAAGGTCTGGAACCT 4451
Db |||||

RESULT 7
AR255867
LOCUS AR255867 5207 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 6482624.
ACCESSION AR255867
VERSION AR255867.1 GI:27305065
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 5207)
AUTHORS Wei, M.-H., Ketchum, K.A., Di Francesco, V. and Beasley, E.M.
TITLE Isolated human kinase proteins, nucleic acid molecules encoding human kinase proteins, and uses thereof
JOURNAL Patent: US 6482624-A 1 19-NOV-2002;
FEATURES
Location/Qualifiers
source 1..5207
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.8%; Score 999.4; DB 6; Length 5207;
Best Local Similarity 99.9%; Pred. No. 8.7e-156;
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCAGTCAGTGACACTGGCTCCAGAGTGTACGCCAGCAGCTGCCAGGCCACCTGG 60
Db 3628 GGCAGTCAGTGACACTGGCTCCAGAGTGTACGCCAGCAGCTGCCAGGCCACCTGG 3687
Qy 61 AGCAAGAGCGAGGCCCTCCAGAGAGCAGCAGCCGTGCTCATCTCTGCACTCAAG 120
Db 3688 AGCAAGAGCGAGGCCCTCCAGAGAGCAGCAGCCGTGCTCATCTCTGCACTCAAG 3747
Qy 121 AACTTCAGCTTCTGACCAATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
Db 3748 AACTTCAGCTTCTGACCAATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3807
Qy 181 AGCGTCAGCAATGCGCTGGGAGCAGTGACCAACCGGGGCTCTCCGGAAGCAGAGCG 240
Db 3808 AGCGTCAGCAATGCGCTGGGAGCAGTGACCAACCGGGGCTCTCCGGAAGCAGAGCG 3867
Qy 241 CCCTCATCTTCGCGCATGCCCGGATATCGGGAGGATGTGTGAGCAAGGAGTGTGCTGGTC 300
Db 3868 CCCTCATCTTCGCGCATGCCCGGATATCGGGAGGATGTGTGAGCAAGGAGTGTGCTGGTC 3927
Qy 301 TGGAGCCCGTGGAAATCTTACCGGCCCTGTGACCTATTGTGCACTAGAGGC 360
Db 3928 TGGAGCCCGTGGAAATCTTACCGGCCCTGTGACCTATTGTGCACTAGAGGC 3987
Qy 361 GGCAGCTGGAACCACTGGCCTCCGACATCTTTGACTGTGCTGCTGCTGCTGCTGCTGCTG 420
Db 3988 GGCAGCTGGAACCACTGGCCTCCGACATCTTTGACTGTGCTGCTGCTGCTGCTGCTGCTG 4047
Qy 421 TCCCGGGTGGCACTTACACCTTCCGCAAGGATGTGTGAGCAAGGAGTGTGCTGGTCCC 480
Db 4048 TCCCGGGTGGCACTTACACCTTCCGCAAGGATGTGTGAGCAAGGAGTGTGCTGGTCCC 4107
Qy 481 TACAGAGCCCTTCGAGCAAGTCTCTCGGAGGGGCCAGCCACCTGCGCTCTGAGGAG 540
Db 4108 TACAGAGCCCTTCGAGCAAGTCTCTCGGAGGGGCCAGCCACCTGCGCTCTGAGGAG 4167
Qy 541 GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTTCGCAATTCAGACA 600
Db 4168 GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTTCGCAATTCAGACA 4227
Qy 601 GAGATCCAGAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTTCGCAATTCAGACA 660
Db 4228 GAGATCCAGAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTTCGCAATTCAGACA 4287
Qy 661 GCGCTGGCGCCCAAGATCATCCCTTACCAACCCAGGACAGACAGTGTGCTGCGCGAA 720
Db 4288 GCGCTGGCGCCCAAGATCATCCCTTACCAACCCAGGACAGACAGTGTGCTGCGCGAA 4347

```
QY 721 TAGAGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCCGAGCTGCACGCGAGCTACCTC 780
DB 4348 TAGAGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCCGAGCTGCACGCGAGCTACCTC 4407
QY 781 AGCCCCGGCACCTGGTGTCTCATCTTGGAGTGTGTCTCTGGGCCCGAGCTGTCTCCCTGC 840
DB 4408 AGCCCCGGCACCTGGTGTCTCATCTTGGAGTGTGTCTCTGGGCCCGAGCTGTCTCCCTGC 4467
QY 841 CTGGCCGAGAGGGCTCTCTACTCAGAACTCCGAGTGAAGGACTACTGTGGCAGATGTTG 900
DB 4468 CTGGCCGAGAGGGCTCTCTACTCAGAACTCCGAGTGAAGGACTACTGTGGCAGATGTTG 4527
QY 901 AGTGCCACCCAGTACTCTCACAAACAGCACATCTCTGCACCTGGACCTCAGGTCCGAGAAC 960
DB 4528 AGTGCCACCCAGTACTCTCACAAACAGCACATCTCTGCACCTGGACCTCAGGTCCGAGAAC 4587
QY 961 ATGATCATCACCGAATACAACTGTCTCAAGTGTCTGGACCT 1001
DB 4588 ATGATCATCACCGAATACAACTGTCTCAAGTGTCTGGACCT 4628

RESULT 8
AR442826
LOCUS AR442826 5207 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 1 from patent US 6670164.
ACCESSION AR442826
VERSION AR442826.1 GI:42670270
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 5207)
Wei, M.-H., Ketchum, K.A., Di Francesco, V. and Beasley, E.M.
TITLE Isolated human kinase proteins
JOURNAL Patent: US 6670164-A 1 30-DEC-2003;
FEATURES
Location/Qualifiers
source
1..5207
/mol_type="genomic DNA"

ORIGIN
Query Match 99.8%; Score 999.4; DB 6; Length 5207;
Best Local Similarity 99.9%; Pred. No. 8.7e-156;
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCCAGTCAGTGACACTGGCTGCAGCTGTGCAGCCAGCAGCTGCCAGGCCACCTGG 60
DB 3628 GGCCAGTCAGTGACACTGGCTGCAGCTGTGCAGCCAGCAGCTGCCAGGCCACCTGG 3687
QY 61 AGCAAGACGAGGCCCTCGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 120
DB 3688 AGCAAGACGAGGCCCTCGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 3747
QY 121 AACTTCCAGCTTCTGACCATCTGTGTGTGTGGCTGAGGACCTGGGTGTGTACACCTGC 180
DB 3748 AACTTCCAGCTTCTGACCATCTGTGTGTGTGGCTGAGGACCTGGGTGTGTACACCTGC 3807
QY 181 AGCGTGAGCAATCGCTGGGACAGTGACACCAACCGCGCTCTCCGGAAGGACAGAGCG 240
DB 3808 AGCGTGAGCAATCGCTGGGACAGTGACACCAACCGCGCTCTCCGGAAGGACAGAGCG 3867
QY 241 CCCTCATCTTCCGCAATGCCGATATCGGGAGAGTGTACGGGATGGGGTGTCTGTGTC 300
DB 3868 CCCTCATCTTCCGCAATGCCGATATCGGGAGAGTGTACGGGATGGGGTGTCTGTGTC 3927
QY 301 TGGAGCCCGTGGAACTCTACGGCCCTGACCTTGTGACATTTGTGCGATGCGAGCTAGAGGC 360
DB 3928 TGGAGCCCGTGGAACTCTACGGCCCTGACCTTGTGACATTTGTGCGATGCGAGCTAGAGGC 3987
QY 361 GCGAGCTGGACACACTGGCTCCGACATCTTTTGAAGTGTCTCTGACCTGACCGAGCAAGTC 420
DB 3988 GCGAGCTGGACACACTGGCTCCGACATCTTTTGAAGTGTCTCTGACCTGACCGAGCAAGTC 4047
```

```
QY 421 TCCCGGGGTGGACCTACACCTTCGGCACGCACTGTGTACGACGAGGAGGANTGGGTCCC 480
DB 4048 TCCCGGGGTGGACCTACACCTTCGGCACGCACTGTGTACGACGAGGAGGANTGGGTCCC 4107
QY 481 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACTTGGCCTCTCAGGAG 540
DB 4108 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACTTGGCCTCTCAGGAG 4167
QY 541 GAGAGCCAGGGGGCGGTGAGCCCAACCTCTGCCAGCACAAGAGACTTTCGCAATCCAGACA 600
DB 4168 GAGAGCCAGGGGGCGGTGAGCCCAACCTCTGCCAGCACAAGAGACTTTCGCAATCCAGACA 4227
QY 601 CAGATCCAGAGGGGCCCTTCCAGCGTGTGGCGCAATGCTGGGAGAGGCCAGCGGGCGG 660
DB 4228 CAGATCCAGAGGGGGCGCTTCCAGCGTGTGGCGCAATGCTGGGAGAGGCCAGCGGGCGG 4287
QY 661 GCGCTGGCGGCCAAGATCATCCCTTACACCCCAAGGACAAAGACAGCAGTGTCTGCGGAA 720
DB 4288 GCGCTGGCGGCCAAGATCATCCCTTACACCCCAAGGACAAAGACAGCAGTGTCTGCGGAA 4347
QY 721 TACGAGGCCCTCAAGGGCTGCGCCACCCGACCTTGGCCAGCTGCGACGAGCTTACCTC 780
DB 4348 TACGAGGCCCTCAAGGGCTGCGCCACCCGACCTTGGCCAGCTGCGACGAGCTTACCTC 4407
QY 781 AGCCCCGGCACCTGGTGTCTCATCTTGGAGCTGTGTCTTGGGCCCGAGCTGTCTCCCTGC 840
DB 4408 AGCCCCGGCACCTGGTGTCTCATCTTGGAGCTGTGTCTTGGGCCCGAGCTGTCTCCCTGC 4467
QY 841 CTGGCCGAGAGGGCTCTCTACTCAGAACTCCGAGTGAAGGACTACTGTGGCAGATGTTG 900
DB 4468 CTGGCCGAGAGGGCTCTCTACTCAGAACTCCGAGTGAAGGACTACTGTGGCAGATGTTG 4527
QY 901 AGTGCCACCCAGTACTCTGCACAAACAGCACATCTCTGCACCTGGAGCTTCCAGGTCGAGAAC 960
DB 4528 AGTGCCACCCAGTACTCTGCACAAACAGCACATCTCTGCACCTGGAGCTTCCAGGTCGAGAAC 4587
QY 961 ATGATCATCACCGAATACAACTGTCTCAAGTGTCTGGACCT 1001
DB 4588 ATGATCATCACCGAATACAACTGTCTCAAGTGTCTGGACCT 4628

RESULT 9
AR442827
LOCUS AR442827 5207 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 3 from patent US 6670164.
ACCESSION AR442827
VERSION AR442827.1 GI:42670271
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 5207)
Wei, M.-H., Ketchum, K.A., Di Francesco, V. and Beasley, E.M.
TITLE Isolated human kinase proteins
JOURNAL Patent: US 6670164-A 3 30-DEC-2003;
FEATURES
Location/Qualifiers
source
1..5207
/mol_type="genomic DNA"

ORIGIN
Query Match 99.8%; Score 999.4; DB 6; Length 5207;
Best Local Similarity 99.9%; Pred. No. 8.7e-156;
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCCAGTCAGTGACACTGGCTGCAGCTGTGCAGCCAGCAGCTGCCAGGCCACCTGG 60
DB 3628 GGCCAGTCAGTGACACTGGCTGCAGCTGTGCAGCCAGCAGCTGCCAGGCCACCTGG 3687
QY 61 AGCAAGACGAGGCCCTCGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 120
DB 3688 AGCAAGACGAGGCCCTCGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 3747
```



```
Db 4468 CTGGCCGAGAGGGCCCTCTACTCAGAACTCTGAGGTGAAGGACTACTCTGTGGCAGATGTTG 4527
Qy 901 AGTGCACCCAGTAGTACCTGCACAAACAGACACATCTCTGCACCTGGAGCTTCCGAGAAC 960
Db 4528 AGTGCACCCAGTAGTACCTGCACAAACAGACACATCTCTGCACCTGGAGCTTCCGAGAAC 4587
Qy 961 ATGATCATCAGCAATACAACTGCTCAAGGTCTGGACCT 1001
Db 4588 ATGATCATCAGCAATACAACTGCTCAAGGTCTGGACCT 4628

RESULT 11
AX430858 AX430858 5207 bp DNA linear PAT 28-JUN-2002
LOCUS Sequence 3 from Patent WO0240683.
DEFINITION AX430858
ACCESSION AX430858
VERSION AX430858.1 GI:21655922
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ketchum, K., Beale, E.M., Wei, M.H. and di Francesco, V.
TITLE Isolated human kinase proteins, nucleic acid molecules encoding
human kinase proteins, and uses thereof
JOURNAL PE CORP NY (US)
FEATURES
    Location/Qualifiers
    source
        1..5207
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match 99.8%; Score 999.4; DB 6; Length 5207;
Best Local Similarity 99.9%; Pred.No. 8.7e-156;
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCAGTCAGTGACACTGGCTGCCAGGTGTGACGCCAGCCAGCTGCGCCAGCCACCTGG 60
Db 3628 GCCAGTCAGTGACACTGGCTGCCAGGTGTGACGCCAGCCAGCTGCGCCAGCCACCTGG 3687

Qy 61 AGCAAGACGAGGCCCCCTCGAGAGCAGCAGCGGTCTCTCATCTTGCACCCCTCAAG 120
Db 3688 AGCAAGACGAGGCCCCCTCGAGAGCAGCAGCGGTCTCTCATCTTGCACCCCTCAAG 3747

Qy 121 AACTTCCAGCTTCTGACCATCTTGGTGTGTGCTGAGGACCTGGGTGTGTACACCTGC 180
Db 3748 AACTTCCAGCTTCTGACCATCTTGGTGTGTGCTGAGGACCTGGGTGTGTACACCTGC 3807

Qy 181 AGCGTGAGCAATGGCTGGGACAGTACACACCGCGCGTCTCCGGAAGCAGAGCGC 240
Db 3808 AGCGTGAGCAATGGCTGGGACAGTACACACCGCGCGTCTCCGGAAGCAGAGCGC 3867

Qy 241 CCCTCATCTTCCGCAATGCCGATATCGGGAGGTGTACCGGATGGGGTGTCTGCTGC 300
Db 3868 CCCTCATCTTCCGCAATGCCGATATCGGGAGGTGTACCGGATGGGGTGTCTGCTGC 3927

Qy 301 TGGAGCCCGTGGAACTCTAGCCCTCTGACCTACATTGTGCAGTCAGCCTAGAAGGC 360
Db 3928 TGGAGCCCGTGGAACTCTAGCCCTCTGACCTACATTGTGCAGTCAGCCTAGAAGGC 3987

Qy 361 GCGAGCTGGACACACTGGCTTCCGACATCTTTGACTCTGCTACTCTACCCAGCAAGCTC 420
Db 3988 GCGAGCTGGACACACTGGCTTCCGACATCTTTGACTCTGCTACTCTACCCAGCAAGCTC 4047

Qy 421 TCCCGGGGTGGCACTACACCTTCCGCAAGCAGCATGTGTGAGCAAGGAGGAGGATGGGTCCC 480
Db 4048 TCCCGGGGTGGCACTACACCTTCCGCAAGCAGCATGTGTGAGCAAGGAGGAGGATGGGTCCC 4107

Qy 481 TACAGCAGCCCTCGGAGCAAGTCTCTCTCGGAGGGCCAGCCACTGGCTCTGAGGAG 540
```

```
Db 4108 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACTGTCCTCTGAGGAG 4167
Qy 541 GAGAGCCAGGGCGGTTCAGCCCAACCCCTGCCCCAGCACAAGAGACTTTCGATTCAGACA 600
Db 4168 GAGAGCCAGGGCGGTTCAGCCCAACCCCTGCCCCAGCACAAGAGACTTTCGATTCAGACA 4227
Qy 601 CAGATCCAGAGGGGCGGTTCAGCGTGTGGCGCAATGCTGGGAGAGGCCAGCGGGCGG 660
Db 4228 CAGATCCAGAGGGGCGGTTCAGCGTGTGGCGCAATGCTGGGAGAGGCCAGCGGGCGG 4287
Qy 661 GCGTGGCGCGCAAGATCATCCCTACACCCCAAGGACAAAGACAGCAGCAGTGTCTGGCGAA 720
Db 4288 GCGTGGCGCGCAAGATCATCCCTACACCCCAAGGACAAAGACAGCAGCAGTGTCTGGCGAA 4347
Qy 721 TACAGAGCCCTCAAGGGCTGCGCCACCCGACCTGCGCCAGCTGCAGCAGCCTACCTC 780
Db 4348 TACAGAGCCCTCAAGGGCTGCGCCACCCGACCTGCGCCAGCTGCAGCAGCCTACCTC 4407
Qy 781 AGCCCCCGGACCTGCTCATCTTGGAGCTGTCTTGGGCGGAGCTGCTTGGGCGGAGCTGCTCCCCTGC 840
Db 4408 AGCCCCCGGACCTGCTCATCTTGGAGCTGTCTTGGGCGGAGCTGCTTGGGCGGAGCTGCTCCCCTGC 4467
Qy 841 CTGGCCGAGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTGGCAGATGTTG 900
Db 4468 CTGGCCGAGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTGGCAGATGTTG 4527
Qy 901 AGTCCACCCAGTAGTACCTGCACAAACAGCAGCATCTCTGCACCTGGAGTCCGAGAAC 960
Db 4528 AGTCCACCCAGTAGTACCTGCACAAACAGCAGCATCTCTGCACCTGGAGTCCGAGAAC 4587
Qy 961 ATGATCATCAGCAATACAACTGCTCAAGGTCTGGACCT 1001
Db 4588 ATGATCATCAGCAATACAACTGCTCAAGGTCTGGACCT 4628

RESULT 12
AX039412 AX039412 7928 bp DNA linear PAT 18-NOV-2000
LOCUS Sequence 5 from Patent WO0063381.
DEFINITION AX039412
ACCESSION AX039412
VERSION AX039412.1 GI:11229480
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zeng, W., Stanton, L. and Kong, H.
TITLE Mammalian protein with putative function in signal transduction
JOURNAL Patent: WO 0063381-A 5 26-OCT-2000;
SCIOS INC. (US)
FEATURES
    Location/Qualifiers
    source
        1..7928
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAC16626.1"
            /db_xref="GI:11229481"
            /translation="MLERPTPKYKKGSSITFSVKVEGRPVPTVHWLREAEABGVLMW
            GPDTPGYTVASSAQHSLVLDVGRHQGTVCTASNAAGALCSASLHVGLPKVEE
            KEVKREALISTFLOGTTTAAISAOGLTASPADLQGRKEEPLAAKALHLSLAEBVT
            EEPLOKLTSTITEMVSAKITQAKLOVPGSDSDSKTPSPSPHGRSPSSSIQSSSS
            ESEGDGAGEIFDIYVYADYLPALQEQDAITLREGQYVEILDAAHPLRWLRVTKPTK
            SSPSGQWSPAYLDRLKUSPEWGAEEPEFGEAVSEDEYKARLSVITQELLUSSEQ
            AFVBEFLQASHHLQHLERCPHPIAVAGQKAVIPRNVRDIFGRHSSFLQBLQCDTD
            DDVACFCIKQAQAEQYLFVGRVQASVVSSTAIQEFYKVAEALPQALPQSQPDP
            PPLQYLEQPVORVYQALLKELIRNARNRONCALLQOAVYVVSALPQAEALKLV
            SLMENYPTLEALGEPIRQGHFIVWEGAPGARMPKGNHVFIPRNHLVFKCPKPRDS
            RTDTSVSVFRNMKLSLIDLDQVEGDRAFEVWQBERDSVRKYLQARTAILKSSWV
```


QY 185 TGAGCAATGCGCTGGGACAGTGACCAACCGGGGCTCTCGGAGGACGAGCGCCCT 244
 Db 17287 TGAGCAATGCGCTGGGACAGTGACCAACCGGGGCTCTCGGAGGACGAGCGCCCT 17346
 QY 245 CATCTTCCCATGCGCGGATATCGGGAGGTGTACCGGATGGGTGCTGCTGCTCGA 304
 Db 17347 CATCTTCCCATGCGCGGATATCGGGAGGTGTACCGGATGGGTGCTGCTGCTCGA 17406
 QY 305 AGCCCGTGGAAATCTACGCGCTGTGACCTATGTCAGTGACGACCTAGAGCGGCA 364
 Db 17407 AGCCCGTGGAAATCTACGCGCTGTGACCTATGTCAGTGACGACCTAGAGCGGCA 17466
 QY 365 GCTGGACCACTGGCTCCGACATCTTTGACTGCTGCTACTGACGACGAGCTCTCC 424
 Db 17467 GCTGGACCACTGGCTCCGACATCTTTGACTGCTGCTACTGACGACGAGCTCTCC 17526
 QY 425 GGGTGGGACCTACACCTTCCGACGCGCATGTGTGACGACGAGGAAATGGTCCCTACA 484
 Db 17527 GGGTGGGACCTACACCTTCCGACGCGCATGTGTGACGACGAGGAAATGGTCCCTACA 17586
 QY 485 CGAGCCCTCGGAGCAAGTCTCTCGGAGGCGCCAGCCACTGGCCCTCTGAGGAGGAGA 544
 Db 17587 CGAGCCCTCGGAGCAAGTCTCTCGGAGGCGCCAGCCACTGGCCCTCTGAGGAGGAGA 17616
 QY 545 GCCAGGCGGTGAGCCCAACCCCTGCCAGCAGCAAAAGACCTTCGCATTCAGACACAGA 604
 Db 17617 GGAATGGAGGGGACCAACCCAGCCCTGACCGCTGTCCCTTGA----- 17661
 QY 605 TCCAGAGGGCGCTTACGCTGTGTGGCGCAATGCTGGGAGAGGAGCCAGCGGGCGGCG 664
 Db 17662 ---GCAGGGCGCTTACGCTGTGTGGCGCAATGCTGGGAGAGGAGCCAGCGGGCGGCG 17718
 QY 665 TGGCGCGCAAGATCATCCCTACCAACCCAGGAGCAAGACAGTGTGCTGCGCGAATACG 724
 Db 17719 TGGCGCGCAAGATCATCCCTACCAACCCAGGAGCAAGACAGTGTGCTGCGCGAATACG 17778
 QY 725 AGGCGCTCAAGGGCTGCGCCACCGGACCTGGCGCCAGCTGACGACGACCTACCTCAGCC 784
 Db 17779 AGGCGCTCAAGGGCTGCGCCACCGGACCTGGCGCCAGCTGACGACGACCTACCTCAGCC 17838
 QY 785 CCGGACCTGCTGCTCATCTTGGAGCTGTGCTGCGGCGGAGCTGCTGCTGCTGCTGCTG 844
 Db 17839 CCGGACCTGCTGCTCATCTTGGAGCTGTGCTGCGGCGGAGCTGCTGCTGCTGCTGCTG 17898
 QY 845 CCGAGAGGCGCTCTACTCAGAAATCCGAGGTGAAGACTACTCTGCGAGATGTTGAGTG 904
 Db 17899 CCGAGAGGCGCTCTACTCAGAAATCCGAGGTGAAGACTACTCTGCGAGATGTTGAGTG 17958
 QY 905 CCACCCAGTACTGACCAACGACGACATCTGACCTGAGCTGAGTCCGAGAACATGA 964
 Db 17959 CCACCCAGTACTGACCAACGACGACATCTGACCTGAGCTGAGTCCGAGAACATGA 18018
 QY 965 TCATCACCAGAAATACACCTGCTCAAGGTCGTTGACCT 1001
 Db 18019 TCATCACCAGAAATACACCTGCTCAAGGTCGTTGACCT 18055

 RESULT 14
 AV603754
 LOCUS Mus musculus Mus musculus 4740 bp mRNA linear ROD 01-AUG-2004
 DEFINITION Mus musculus obscurin-MLCK mRNA, complete cds.
 ACCESSION AY603754
 VERSION AY603754.1 GI:50660397
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 4740)
 Sutter, S.B., Raeker, M.O., Borisov, A.B. and Russell, M.W.
 Orthologous relationship of obscurin and Unc-89: phylogeny of a
 novel family of tandem myosin light chain kinases

JOURNAL Dev. Genes Evol. 214 (7), 352-359 (2004)
 PUBMED 15185077
 REFERENCE 2 (bases 1 to 4740)
 AUTHORS Sutter, S.B., Raeker, M.O. and Russell, M.W.
 TITLE Direct Submission
 JOURNAL Submitted (21-APR-2004) Pediatric Cardiology, University of
 Michigan, 1500 E. Medical Center Dr., Ann Arbor, MI 48109, USA
 FEATURES Location/Qualifiers
 source
 1..4740
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 CDS
 1..4740
 /codon_start=1
 /product="obscurin-MLCK"
 /protein_id="AAT80900.1"
 /db_xref="GI:50660398"
 /translation="MOVITIEDVQVQDMAQFDAVIEGHPPPIVTWYKSGTOLTSAR
 LSOQDCTTVSLVLTVDVPHDAGVYTCVANNAGGOVLCKAEALLVHGDKLDAENQVYR
 KKLHSFDVBEIEGRGVFGFVKRVQHQNGNMFCAKFIPLRSKTRAQYQBRDLATL
 GHLVTLGLOQFETRLKLLILELCSSEELDLRFKGVVTEAEVKYIQQLVGGLHY
 LHSGLIHLDDIKPENILMVHAREDIKIDFGFAKITPSEPQSKYSGVSPVAFPEII
 EQNPVSGSDIWMGVISYLSLTCSPPAGESDRATLLNVLEGVSSWSPFAHLS
 AKDFIKATLQRTPOARSTOCIAHPWFLKMPAEAEHFINTKQLKLARWQSL
 MSYKSIIVMSIPELLOQPSLGVARHLRGEASGASSSSSSSSSSSSSSSSSSSSSS
 PPSVTHSLPHRPFRLPSASLPEETEASMTADAAVPASQAPGAPSPASGCVPRHS
 VISLFFYQQAEGAEKNTSGAKRHPARRHLKGGYIARALPGLRPLMEYLSLEE
 EAAEQAQSLMTKTPFETALRLPSSVREVPGRSHSLNDNPVTTGPPSPACKQOLLF
 PPSTGLTHETAKDRGHEGFLQSPVFPMSGDSRPGKQSGSDSCRGKPAAGSCHS
 ELGSGGEGGPPSSQSLGSLPQSLKELSTSCGPLFSEQPAAPPTVPSPLGSE
 KEDQSGLSGSPVPSSSPGSASQVDSLDTGSLSEAGDTCDTTPPQRPQOATTR
 KPSLESRGYAGVAGYGTAFGGDAGMCGPLWARMANAVSSGSEEDQANATESPQ
 PLESGLPIASGVPLRTPSLTPWEVEQVSLVQIRLDSGDAADTISLDSISVDP
 AYNLSDLYDIKYLFPFEMI FRVPKPIEQPESGSETEAGQGLADFLLEAAWMPGE
 LGRLAGEIITEPEEPDGLLEAGEAAVGRKRKWSRGLFQFPGRCDSLSEPEVLGL
 RORVKAEMHISRLIKGRPEGPRPKAGLASRLSLKGRDLKGRDLSDEAVVLGO
 SVTLACOVLAQPTAQTMSKDVLLSESHLLISLTKNFKOLLTLVVKPESGPTVYIQQCCI
 CYNPLGTAVTTGVLRKAERSSSPREVEGELYKADVLLWKPVESGPTVYIQQCCI
 EBSGWTTLASDISDCYLTGKLSRGMYIFRTACVSKAGMGPPSPSEQVJLGGPNHL
 ASBESRGRPAQLLPSTKTFAPQIIRGRFVVRQEKASGRALAAKATVPYQPED
 KTVLREYKALKRLHPHLAAHLYSLRHLVILRLCSGPELLSPSLAESYSESD
 VKDYLMQMSATQVYLAHILHLDRSENMMVTBYNLLKVIDLGNQSLDQKVPAPPE
 NKDYLETMAPELLEGGQAVPOTDIWAIGVTAFIMLSEVPESSEGTDRDLQKGRKL
 IRLSRVAGLSGGNAVFLQSLCAQPWGRPCASTCLOGGMLTEBEGTSGSRTPPTFT
 VLRLAFVREKRRALLYKKHNLAAQR"
 439..1206
 /note="Region: serine/threonine protein kinase domain"
 4015..4725
 /note="Region: serine/threonine protein kinase domain"
 ORIGIN
 Query Match 73.8%; Score 738.4; DB 10; Length 4740;
 Best Local Similarity 84.2%; Pred. No. 1.3e-112;
 Matches 845; Conservative 0; Mismatches 156; Indels 3; Gaps 1;
 QY 1 GGCAGTCAGTGACACTGGCTGCCAGTGTGACCCAGCCAGCTGCCAGCCACCTGG 60
 Db 3259 GGCCAATCAGTGACACTGGCTGCCAGTGTGACCCAGCCAGCTGCCAGCTGG 3318
 QY 61 AGCAAGACGAGCGCCCTGGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 120
 Db 3319 AGCAAGATGGGTGCTCTGGAGAGCAGCGCCACCTCTCTCATCTCTCCACCTGAAG 3378
 QY 121 AACTTCCAGCTTCTGACCATCTCTGGTGGTGGTCTGAGGACCTGGGTGTGTACACCTGC 180
 Db 3379 AACTTCCAGCTTCTGACCATCTCTGGTGAAGAGGAGGATCTGGGCACATATACCTGC 3438
 QY 181 AGCTGAGCAATGCGCTGGGACAGTGAACACAGCGGCTGCTCCCGAAGAGCAGAGCGC 240
 Db 3439 TGTGTGACCAACCGCTGGGACAGCAGTCAACACAGGTGTCTCCCGAAGAGCAGAGCGC 3498
 QY 241 CCCTCATCTTCGCGATCGCGGATATCGGAGAGGTGTACCGGATGGGTCTCTGCTGTC 300

```
Db 3499 CCTTCATCTTCTCCACGCCCTGAGGTGGGGAACTATACAAGGATGGGTGCTGCTGGTC 3558
Qy 301 TGAAGGCCGTGAATCTACGCGCTGTGACCTACATATGTCAGTGCAGTGCCTAGAGCC 360
Db 3559 TGGAGCCTGTGAATCTGTGCGCGTGGACCTACATATGTCAGTGCCTAGAGGAG 3618
Qy 361 GGCAGCTGGACACATCGGCTCCGACATCTTTGACTGTGCTGCTACCTGACCAAGCAAGCTC 420
Db 3619 GGCAGCTGGACAAACCTCGCTTCCAGACATCTCTGACTGTGCTACCTGCTGCTGCAAGCTA 3678
Qy 421 TCCCGGGGTGGACCTACACCTTCGCGACGGCATGTGTGAGCAGGAGGAGGAGTGGTCCC 480
Db 3679 TCGAGGGGTGGCATGTATATCTTCCGGACGAGCATGTGTGAGCAAGCAGGAATGGGCCCC 3738
Qy 481 TACAGCAGGCCCTCGGAGCAAGTCTCTCGGAGGGCCCGACCATCGCTGCGCTCAGGAG 540
Db 3739 TACAGCAGGCCCTCAGAACAGGTCTCTCTCGGAGGACCCCAACCATCTGCGCTCAGGAG 3798
Qy 541 GAG---AGCCAGGGCGGTGAGCCCAACCCCTGCCAGCACAAAGACCTTGCATTCCAG 597
Db 3799 GAAAGCAGCGGGGGAGGCGAGCCAGCTTCTCCCGAGCACAAAGACCTTTGSCCTTCCAG 3858
Qy 598 ACACAGATCCAGAGGGCGCTTACGCTGTGTGCGGCAATGCTGGAGAGGCCAGCGGG 657
Db 3859 ATGCAGATCCGAGGGCGCGCTTACGTGTGTGCGGCAATGCTGGAGAGGCCAGAGTGG 3918
Qy 658 CGGGCGCTGGCGCCCAAGATATCCCTACCAACCCCAAGGACCAAGACAGCAGTGTGCGC 717
Db 3919 CGGGCGCTAGCTGTGAATCTTCTTACCAACCTGAGGACAGCAGCTGTACTAGA 3978
Qy 718 GAATAGAGGCCCTCAAGGGCTGCGCCACCGCACCTGCGCCAGCTGCGACGCTTAC 777
Db 3979 GAATAGAGGCCCTTAAAGAGACTGACCAACCCACATCTGGCCCACTCCATCGCGCTTAC 4038
Qy 778 CTCAGCCCCCGGCACCTGTGTCTCATCTTGGAGCTGTCTCTGGCCCGAGCTGTCCCC 837
Db 4039 CTCAGTCCCGGCACCTGTGTCTCATCTTGGAGCTGTCTCTGGCCCGAGCTGTACCC 4098
Qy 838 TCCTGGCGGAGAGGGCCCTCTACTCAGAACTCCGAGGTGAAGGACTACTGTGGCAGATG 897
Db 4099 TCTTTGGCAGAGAGGGAATCTCTACTCAGAGTCTGATGTGAAGGACTACTGTGGCAGATG 4158
Qy 898 TTGAGTGCACCCAGTACTGACACACAGCAGACATCTTGCACCTGAGCTGAGTCCGAG 957
Db 4159 CTGAGTGCACCCAGTACTGATGCCCAACACATCTTGCACCTGAGCTGAGTCCGAG 4218
Qy 958 AACATGATCATCACCAGAAATAACCTGCTCAAGGTCTGAGGACCT 1001
Db 4219 AACATGATGCTCACTGAGTACAACTGCTTAAGGTATAGACCT 4262

RESULT 15
BC046431 1183 bp mRNA linear ROD 16-SEP-2003
LOCUS Mus musculus cDNA sequence BC046431, mRNA (cDNA clone
IMAGE:4006702), partial cds.
ACCESSION BC046431
VERSION BC046431.1 GI:28386252
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1183)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schnee, T.E., Brownstein, M.J., Udén, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
```

```
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Jones, S.J., Skalska, U., Smalios, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477332
2 (bases 1 to 1183)
Straussberg, R.
Direct Submission
Submitted (03-FEB-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lousseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 92 Row: d Column: 18.
FEATURES
Location/Qualifiers
1..1183
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:4006702"
/tissue_type="Mammary tumor metastasized to lung.
MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR
enhancer."
/clone_lib="NCI CGAP_Lu30"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
<1..1183
/gene="BC046431"
/db_xref="LocusID:216790"
/db_xref="MGI:2664530"
<1..1027
/feature="MGI:2664530"
/codon_start=2
/product="BC046431 protein"
/protein_id="AAH46431.1"
/db_xref="GI:28386253"
/db_xref="LocusID:216790"
/db_xref="MGI:2664530"
/translation="VSKAGMGYPSPSQVLLGGPNHLASEESRGPQALLPSTYKT
FAFQMIQEPGSPVQRCKRAXSALAAKIVPQPEKTVLEIEAKRLHHPLHA
QLHAYLSPHMLVLELDCSGPELLPSLAERESYSDVDKYLQMLSATQYLAHOHI
LHLDUSRNHVVLELVDLGNQSLDQEKVPAPENFKLIRLRCVAGLGGAVFLQS
PQTDAIAGTAFIMLSGEYPSSEGTDLQKLRKGLIRLRCVAGLGGAVFLQS
SLCQAPWGRPCASTCLOCGMLTEBPTGSRPTPTVPTVRLRAFVREKRRLYYK
HNLQAVR"
misc_feature
134..892
/feature="BC046431"
```

/note="S TKc; Region: Serine/Threonine protein kinases,
catalytic_domain"
/db_xref="CDD:smart00220"

ORIGIN

```
Query Match      40.6%; Score 406.6; DB 10; Length 1183;
Best Local Similarity 85.1%; Pred. No. 1.3e-57;
Matches 467; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

Qy 456 TGTACGAAAGCAGGAATGGTCCCTACAGAGCCCTCGAGCAAGTCTCTCGGAGG 515
Db      |||||
1 TGTACGAAAGCAGGAATGGGCCCCCTACAGAGCCCTCAGAAACAGTCTCTCGGAGG 60

Qy 516 GCCAGCCACTGGCCTGTAGGAGGA---GAGCCAGGGGGCGTCAGCCCAACCCCTGCC 572
Db      |||||
61 ACCCAACACCTGGCCTGTAGGAGGAAGCAGCGGGGGAGGCCAGCCCAAGTCTCTCCC 120

Qy 573 CAGCAAAAGACCTTTCGATTCACACACAGATCCAGAGGGGCCGCTTCAGCGTGGTGG 632
Db      |||||
121 CAGCAAAAGACCTTTCGCTTCAGATCCGAGGGGCCGCTTCAGTGGTGG 180

Qy 633 GCAATGCTGGGAGAAAGCCAGCGGGCGCTGGCCGCCAAGATCATCCCTACCAACC 692
Db      |||||
181 GCAGTGCAGGAGAAAGCAAGTGGGGGCCCTAGCTGCTAAGATCGTTCCTTACCAACC 240

Qy 693 CAAGCACAAGACAGCTGTGCTGCGCGAATACGAGGCCCTCAAGGGCCTGCGCCACCCGCA 752
Db      |||||
241 TGAGGACAAGACAGCTGTACTAAGAGAAATACGAGGCCTTAAGAGACTGCACCAACCACA 300

Qy 753 CCTGCCCCAGCTGCACGAGCCTACTCAGCCCGCGCACCTGGTGTCTCATCTTGGAGCT 812
Db      |||||
301 TCTGCCCCAACTCCATGCGGCTACTCAGTCCCCGGCACCTGGTGTCTCATCTTGGAGCT 360

Qy 813 GTGCTCTGGGCCCGAGCTGCTCCCTGGCTGGCCGAGAGGGCCTCCTACTCAGAAATCCGA 872
Db      |||||
361 GTGCTCTGGCCCTGAGCTGCTACCCCTTTGGCGGAGAGGAATCCTACTCAGAGTCTGA 420

Qy 873 GGTGAAGGACTACTGTGGCAGATGTTGAGTGCCACCCAGTACCTGCGCAACCAACACAT 932
Db      |||||
421 TGTGAAGGACTACTGTGGCAGATGCTGAGTGCCACCCAGTACTTTGCAATGCCCAACACAT 480

Qy 933 CCTGCACCTGGACCTGAGTCCGAGAACATGATCATACCGGAATACAACCTGCTCAAGT 992
Db      |||||
481 CCTGCACCTGGACCTGAGTCCGAGAACATGATGCTCACTGAGTACAACCTGCTTAAGT 540

Qy 993 CGTGGACCT 1001
Db      |||||
541 TATAGACCT 549
```

Search completed: March 21, 2005, 05:34:55
Job time : 4521.62 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:24:10 ; Search time 4515.62 Seconds
(without alignments)
10741.310 Million cell updates/sec

Title: US-10-077-130-4_COPY_3120_4120
Perfect score: 1001
Sequence: 1 aaagatgggaagagctgag.....tcagagcccaaggcggtgtt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	20435	9	HSJ2535 Homo sapi
2	775.8	77.5	18524	6	CQ730656 Sequence
3	279.4	27.9	19475	9	HSJ314900 Homo sapi
4	279.4	27.9	135964	9	AL353593 Human DNA
5	248.4	24.8	2768	6	AR542178 Sequence
6	218.6	21.8	2488	6	BD160445 Primer fo
7	218.6	21.8	2488	6	AX883644 Sequence
8	218.6	21.8	2488	9	AK024186 Homo sapi
9	218.6	21.8	24545	9	HSJ314898 Homo sapi
10	218.6	21.8	174612	2	AC023889 Homo sapi
11	218	21.8	3956	6	CQ851220 Sequence
12	218	21.8	3956	9	AK128447 Homo sapi
13	216.4	21.6	6400	9	HSJ314903 Homo sapi
14	215.4	21.5	10091	9	HSJ314901 Homo sapi
15	207.6	20.7	164766	2	AC026657 Homo sapi
16	207.6	20.7	211829	10	AL645854 Mouse DNA
17	207.4	20.7	260998	2	AC099089 Rattus no
18	202.6	20.2	244553	2	AC098133 Rattus no
19	201	20.1	44358	2	AC142478 Rattus no

C	20	180.2	18.0	25087	2	AC096931	AC096931 Rattus no
	21	140.8	14.1	5314	10	AK122340	AK122340 Mus muscu
	22	140.2	14.0	5378	6	CQ875354	CQ875354 Sequence
	23	140.2	14.0	5382	9	AB014557	AB014557 Homo sapi
	24	137.6	13.7	5783	6	CQ715191	CQ715191 Sequence
C	25	136.8	13.7	526	11	BV077840	BV077840 S208P6660
	26	123.2	12.3	1704	9	AF035292	AF035292 Homo sapi
	27	123.2	12.3	3316	9	BC007201	BC007201 Homo sapi
	28	123	12.3	2184	6	BD160248	BD160248 Primer fo
	29	123	12.3	2184	6	AX883309	AX883309 Sequence
	30	123	12.3	2184	9	AK023854	AK023854 Homo sapi
	31	120	12.0	2025	9	AK025946	AK025946 Homo sapi
	32	108.8	10.9	3015	10	BC056205	BC056205 Mus muscu
	33	104	10.4	724	6	BD150413	BD150413 Primer fo
	34	104	10.4	724	6	AX870351	AX870351 Sequence
C	35	103.4	10.3	168144	2	CR792456	CR792456 Danio rer
	36	102.2	10.2	2417	10	BC051643	BC051643 Mus muscu
	37	100.4	10.0	933	9	BC061909	BC061909 Homo sapi
	38	97.4	9.7	2761	9	BC043916	BC043916 Homo sapi
	39	95.8	9.6	2534	6	AX834822	AX834822 Sequence
	40	95.8	9.6	2534	9	AK097489	AK097489 Homo sapi
	41	80.2	8.0	195690	10	AC098726	AC098726 Mus muscu
	42	80	8.0	3969	5	BC046871	BC046871 Xenopus l
C	43	73.2	7.3	193064	2	AC149860	AC149860 Papio anu
C	44	73.2	7.3	205372	2	AC150609	AC150609 Callithri
C	45	72.6	7.3	208880	10	AC123837	AC123837 Mus muscu

ALIGNMENTS

RESULT 1	HSJ2535	Homo sapiens mRNA for obscurin (OBSCN gene).	20435 bp	mRNA	linear	PRI 14-SEP-2001
LOCUS	HSJ2535	Homo sapiens mRNA for obscurin (OBSCN gene).	20435 bp	mRNA	linear	PRI 14-SEP-2001
DEFINITION	HSJ2535	Homo sapiens mRNA for obscurin (OBSCN gene).	20435 bp	mRNA	linear	PRI 14-SEP-2001
ACCESSION	AJ002535	AJ002535.1 GI:15026973				
VERSION	AJ002535.1	GI:15026973				
KEYWORDS	OBSCN gene; obscurin.					
SOURCE	Homo sapiens (human).					
ORGANISM	Homo sapiens					
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Young, P., Ehler, E. and Gautel, M.					
TITLE	Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomeric assembly					
JOURNAL	J. Cell Biol. 154 (1), 123-136 (2001)					
MEDLINE	21342081					
PMID	11448995					
REFERENCE	2	(bases 1 to 20435)				
AUTHORS	Gautel, M.S.					
TITLE	Direct Submission					
JOURNAL	Submitted (31-OCT-1997) Gautel M.S., Structural Biology Division, European Molecular Biology Laboratory, Meierhofstr. 1, Heidelberg, 69117, GERMANY					
REMARK	Revised by author 20-JUL-2001					
FEATURES	Location/Qualifiers					
source	1..20435					
	/organism="Homo sapiens"					
	/mol_type="mRNA"					
	/db_xref="taxon:9606"					
	/chromosome="1"					
	/map="1q42.1"					
	/cell_type="cardiomyocyte"					
	/tissue_type="cardiac"					
	/dev_stage="adult"					
gene	1..20435					
	/gene="OBSCN"					
CDS	72..19934					
	/gene="OBSCN"					
	/function="titin binding"					
	/codon_start=1					
	/evidence=experimental					


```
QY 301 AGCTCAAAAGTGGGATGGAGTCAAAAGGTGTCACACGAGGAGCTGGTGTGTCACAGGGC 360
Db 3420 AGCTCAAAAGTGGGATGGAGTCAAAAGGTGTCACACGAGGAGCTGGTGTGTCACAGGGC 3479
QY 361 GCGAAAGCAGATGCTGGGAGTACAGCTGTGAGGCTGGGGGACAGAGTCTCTTCCAC 420
Db 3480 GCGAAAGCAGATGCTGGGAGTACAGCTGTGAGGCTGGGGGACAGAGTCTCTTCCAC 3539
QY 421 CTGCACATCAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATTAATGAGTG 480
Db 3540 CTGCACATCAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATTAATGAGTG 3599
QY 481 CAGGCTGAGGCGGGACCACTGCCATGCTGAGCTGTGAGGTGGCCACGCCCCAGAGCGAG 540
Db 3600 CAGGCTGAGGCGGGACCACTGCCATGCTGAGCTGTGAGGTGGCCACGCCCCAGAGCGAG 3659
QY 541 GTGAGCTGTACAGGACGGGAAGAGTGTGAGCTCCAGCTCAAAAGTACGATGAGGTGTC 600
Db 3660 GTGAGCTGTACAGGACGGGAAGAGTGTGAGCTCCAGCTCAAAAGTACGATGAGGTGTC 3719
QY 601 AAGGCTGCACAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATTAATGAGTG 660
Db 3720 AAGGCTGCACAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATTAATGAGTG 3779
QY 661 AGCTGCGAGGCTGGGGGACAGAGTCTCTTTTCAACTGCACATCACAGAGCCCAAGGCA 720
Db 3780 AGCTGCGAGGCTGGGGGACAGAGTCTCTTTTCAACTGCACATCACAGAGCCCAAGGCA 3839
QY 721 GTGTTTGCAGAGGACAGTGTGTGATTAATGAGGTGCGGACTGAGGCGAGGGCCAGTGCC 780
Db 3840 GTGTTTGCAGAGGACAGTGTGTGATTAATGAGGTGCGGACTGAGGCGAGGGCCAGTGCC 3899
QY 781 ACATGAGCTGTGAGGTGCCCGCCAGAGGAGTCAAGTGCATTAAGGAGTGGGAAG 840
Db 3900 ACATGAGCTGTGAGGTGCCCGCCAGAGGAGTCAAGTGCATTAAGGAGTGGGAAG 3959
QY 841 AAGTGAAGTCCAGTTCGAAAGTGGCATAGAGTGGGGTGGCATGGGCGAGTGGTG 900
Db 3960 AAGTGAAGTCCAGTTCGAAAGTGGCATAGAGTGGGGTGGCATGGGCGAGTGGTG 4019
QY 901 GTGCAGAGGAGGCGAGGAGTCTGGGAGTACACTGTGAGGCTGGGGGCGCAGCGG 960
Db 4020 GTGCAGAGGAGGCGAGGAGTCTGGGAGTACACTGTGAGGCTGGGGGCGCAGCGG 4079
QY 961 CTCTCTTCCACTGGATGTTTCAGAGCCCAAGGCGGTGT 1001
Db 4080 CTCTCTTCCACTGGATGTTTCAGAGCCCAAGGCGGTGT 4120

RESULT 2
CQ730656 18524 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 16590 from Patent WO02068579.
DEFINITION CQ730656
ACCESSION CQ730656
VERSION CQ730656.1 GI:42305092
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,B.W.
TITLE Kite, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 16590 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1..18524
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
```

```
Query Match 77.5%; Score 775.8; DB 6; Length 18524;
Best Local Similarity 86.6%; Pred. No. 1.2e-146;
Matches 855; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1 AAAGATGGGAAGAGCTGAGCTCCAGCTTGAAGTGCATGTAGAGGCCAAAGGCTGCAGA 60
Db 981 AAGGATGGGAAGAGCTGAGCTCCAGCTTCGAAAGTGTGATGAGAGGCCACAGCTGCAGC 1040
QY 61 CGAGGCTGGTGGTGCAGCAGGCGCAAGACGATGCCGGGAGCTACAGCTGCCAGGCC 120
Db 1041 CGAGGCTGGTGGTGCAGCAGGCGCGAGCGATGCCGGGAGCTATAGCTGCCAGGCT 1100
QY 121 AGGGCCAGAGGGTCTCTTCCGCTGCACATCACAGAGCCCAAGATGATGTTGCAAG 180
Db 1101 GGGGCCAGAGGGTCTCTTCCATCTGGATGTCAAAGAGCCCAAGTGTGTTTCCAG 1160
QY 181 GAGCAGTCAGTGCATAATGAGGTGCAGGCTGAGGGCGGGCCAGTGCCATCTGAGCTGT 240
Db 1161 GACCAAGTGGCACACAGTGTGAGGTGAGGCGGGGCCAGTGCCACGCTGAGCTGC 1220
QY 241 GAGGTGGCCAGGCCCAGACGAGGTGACGTGGTACAAGAGTGGGAAGAGCTGAGCTCC 300
Db 1221 GAGGTGGCCAGGCCCAGACGAGGTGATGTGTTACAAGATGGGAAGAGCTGAGCTCC 1280
QY 301 AGCTCAAAAGTGGGCATGGAGGTCAAAGGTGCACAACGAGGCTGGTGTGCCACAGGG 360
Db 1281 AGCTTGAAGTGCATGTAGAGGCCAAAGGCTGCAGACGAGGCTGGTGTGCAGCAGGCA 1340
QY 361 GGCAGCAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGGCCAGAGAGTCTCTTCCAC 420
Db 1341 GGCAGATCGGATGCCGGGACTACAGCTGCCAGGCGCAGGGGCTCTCTTCCGC 1400
QY 421 CTGCACATCACAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATAATGAGTG 480
Db 1401 CTGCACATCACAGAGCCCAAGATGATGTTTGCAGAGGAGCAGTCAAGTGCATAATGAGTG 1460
QY 481 CAGGCTGAGGCGGGACCACTGCCATGCTGAGCTGTGAGGTGGCCAGGCCACAGCGAG 540
Db 1461 CAGGCTGAGGCGGGGCCAGTGCCATGCTGAGCTGTGAGGTGGCCAGGCCACAGCGAG 1520
QY 541 GTGAGCTGTACAAGGACGGGAAGCTGAGTCCAGCTCAAAAGTACGATGAGGTC 600
Db 1521 GTGAGCTGTACAAGGATGGGAAGAGCTGAGCTCCAGCTCAAAAGTGGGCATGAGGTC 1580
QY 601 AAGGCTGCACAAGAGGCTGTGATGCAGCAGAGTGGGCAAGCAGAGTCTCGGGAGTAC 660
Db 1581 AAAGGCTGCACAAGAGGCTGTGTCGCCACAGAGCGGGCAAGCAGATGCTGGGGAGTAC 1640
QY 661 AGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTTCAACTGCACATCACAGAGCCCAAGGCA 720
Db 1641 AGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTTCAACTGCACATCACAGAGCCCAAGG 1700
QY 721 GTGTTTGCAGAGGAGCAGTGTGTCATAAAGAGTGGGAGTGGAGGCGGCGGAGTGGC 780
Db 1701 GTGTTTGCAGAGGAGCAGTGTGTCATAAAGAGTGGGAGTGGAGGCGGCGGAGTGGC 1760
QY 781 ACATGAGCTGTGAGGTGGCCCGCCAGAGGAGTCAAGTGCATTAAGGATGGGAAG 840
Db 1761 ATGCTGAGCTGTGAGGTGGCCCGCCAGAGGAGTGTGATGAGTCAAGAGCGGGAAG 1820
QY 841 AAGTGAAGTCCAGTTCGAAAGTGGCATAGAGGCTGGGGTGCATGCGGCGAGCTGGTG 900
Db 1821 AAGTGAAGTCCAGTTCGAAAGTGGCATAGAGGCTCAAGAGGCTGCACACGAGGCTGGTA 1880
QY 901 GTGCAGAGGCGAGGCCAGGAGTCTGGGAGTACACTGTGAGGCTGGGGGCCAGCGG 960
Db 1881 GTGCAGAGGCTGGGCAAGAGCAGATGCTGGGAGTACAGTGTGAGGCTGGGGGCCAGAGA 1940
QY 961 CTCTCTTCCACTGGATGTTTTCAGAG 987
Db 1941 GTCTCTTTCAACTGCACATCACAGAG 1967
```

```
RESULT 3
HSA314900      19475 bp      DNA      linear      PRI 21-MAY-2002
LOCUS          Homo sapiens partial OBSCN gene for obscurin, exons all-a16.
DEFINITION
ACCESSION      AJ314900
VERSION        AJ314900.1 GI:21104329
KEYWORDS       OBSCN gene; obscurin.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS        Young, P., Ehler, E. and Gautel, M.
TITLE          Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere
               assembly
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 19475)
AUTHORS        Gautel, M.S.
TITLE          Direct Submission
JOURNAL        Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie,
               Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse
               11, Dortmund, 44227, GERMANY
FEATURES       source
               1..19475
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               4472..18215
               /gene="OBSCN"
               join(<4472..4747,5855..5860,6702..6977,7817..8092,
               11254..11529,17940..>18215)
               /gene="OBSCN"
               join(<4472..4747,5855..5860,6702..6977,7817..8092,
               11254..11529,17940..>18215)
               /gene="OBSCN"
               /codon_start=3
               /product="obscurin"
               /protein_id="CAC85747.1"
               /db_xref="GI:21104330"
               /db_xref="UniProt/TREMBL:Q8NH7"
               /translation="PKVVFADQVAHSEVQAEASATLSCVEVAQAQTEVMWYKDGKK
               LSSLVKVAEAGKRRLLVVQQAQKTDAGYSCARGQSVFRHLHTEPKRMVFAKQDS
               VHNVEQAEAGASMLSCVEVAQAQTEVTWYKDKLSSSKVGMVEKGTRLVLVPAQ
               KADAGEYCEAGGQSVFHLHTEPKGVFAKEQSVHNEVQAEAGTAMLSCVEAQPOT
               EWTYKDGKLLSSSKVMEVKGCTRLVVQQAQKADAGEYSCARGQSVFOLHTE
               PKAVFABQLVHNEVTRTAGASATLSCVEVAQAQTEVTWYKDKLSSSKVRIEAGC
               MROLVQQAQADAGEYTCGAGGQRLSPHLVSEPKAVFAKEQLAHRKVAQEAIAIT
               LSCVEVAQAQTEVTWYKDKLSSSKVMEAVGCTRLVVQQAQADTGEYSCGAGQ
               RLSSLVAEPKVPFAKEQPVHREVVQQAQAGASTLSCVEVAQAQTEVMWYKDGKKLSFS
               SKVRMEAVGCTRLVVQQAQAVAGEYSCAGSQRLSFHLHVA"
               <4472..4747
               /gene="OBSCN"
               /note="a11"
               4748..5584
               /gene="OBSCN"
               /note="a11"
               5585..5860
               /gene="OBSCN"
               /note="a12"
               5861..6701
               /gene="OBSCN"
               /note="a12"
               6702..6977
               /gene="OBSCN"
               /note="a13"
               6978..7816
               /gene="OBSCN"
               /note="a13"
               7817..8092
               /gene="OBSCN"
               /note="a14"
               8093..11253
```

```

/ gene="OBSCN"
/ note="a14"
11254..11529
/ gene="OBSCN"
/ note="a15"
11530..17939
/ gene="OBSCN"
/ note="a15"
17940..>18215
/ gene="OBSCN"
/ note="a16"

exon
intron
exon
ORIGIN
Query Match      27,9%; Score 279,4; DB 9; Length 19475;
Best Local Similarity 99,6%; Pred. No. 1.8e-46;
Matches 280; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 153 CACAGAGCCCAAGATGATGTTTGCAAAAGGAGCAGTCAGTCATATGAGGTGCAGGCTGA 212
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
5580 CTCAGAGCCCAAGATGATGTTTGCAAAAGGAGCAGTCAGTCATATGAGGTGCAGGCTGA 5639
QY 213 GGCAGGGGCCAGTCGCCATGCTGAGCTGTGAGGTGGCCAGCCAGCCGAGGTGACGTG 272
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
5640 GGCAGGGGCCAGTCGCCATGCTGAGCTGTGAGGTGGCCAGCCAGCCGAGGTGACGTG 5699
QY 273 GTACAAGGATGGGAAGAAGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGGTCAAAAGGGTG 332
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
5700 GTACAAGATGGGAAGAAGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGGTCAAAAGGGTG 5759
QY 333 CACACGGAGGCTGCTGCTGCCACAGCGGGGCAAGCAGATGCTGGGAGTACAGCTGTGA 392
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
5760 CACACGGAGGCTGCTGCTGCCACAGCGGGGCAAGCAGATGCTGGGAGTACAGCTGTGA 5819
QY 393 GGCTGGGGGCCAGAGAGTCTCTTCCACCTGCACATCACAG 433
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
5820 GGCTGGGGGCCAGAGAGTCTCTTCCACCTGCACATCACAG 5860
```

```
RESULT 4
LOCUS          AL353593
DEFINITION     Human DNA sequence from clone RP5-1139B12 on chromosome 1q42.1-43,
               complete sequence.
ACCESSION      AL353593
VERSION        AL353593.33 GI:18673899
KEYWORDS       HTG.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
               Almeida, J.
REFERENCE      1 (bases 1 to 135964)
AUTHORS        Direct Submission
TITLE          Submitted (13-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
               Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
               humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
               On Feb 14, 2002 this sequence version replaced gi:17977879.
               During sequence assembly data is compared from overlapping clones.
               Where differences are found these are annotated as variations
               together with a note of the overlapping clone name. Note that the
               variation annotation may not be found in the sequence submission
               corresponding to the overlapping clone, as we submit sequences with
               only a small overlap as described above.
               This sequence was finished as follows unless otherwise noted: all
               regions were either double-stranded or sequenced with an alternate
               chemistry or covered by high quality data (i.e., phred quality >=
               30); an attempt was made to resolve all sequencing problems, such
               as compressions and repeats; all regions were covered by at least
               one plasmid subclone or more than one M13 subclone; and the
               assembly was confirmed by restriction digest. The following
               abbreviations are used to associate primary accession numbers given
               in the feature table with their source databases: Em:, EMBL; Sw:,
               SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; information on the WORMPEP
```

database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
 RP5-1139B12 is from the library RPi-5 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP5-1139B12.

FEATURES

source

1. 135964

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="1"

/map="q42.1-43"

/clone="RP5-1139B12"

/clone_lib="RPi-5"

misc_feature

25928..26054
 /note="Sequence from overlapping clone RP11-520H14
 (AL359510). Assembly confirmed by restriction digest."

ORIGIN

Query Match 27.9%; Score 279.4; DB 9; Length 135964;
 Best Local Similarity 99.6%; Pred. No. 1.3e-46;
 Matches 280; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 153 CACAGAGCCCAAGATGATGTTTGCAGAGGAGCAGTCAGTGCATATGAGGTGCAGGCTGA 212
 DB 31618 CTCAGAGCCCAAGATGATGTTTGCAGAGGAGCAGTCAGTGCATATGAGGTGCAGGCTGA 31677
 QY 213 GCGGGGGCCAGTGCATGCTGAGTGTGAGGTGGCCAGGCCCAAGAGGTGACGTG 272
 DB 31678 GCGGGGGCCAGTGCATGCTGAGTGTGAGGTGGCCAGGCCCAAGAGGTGACGTG 31737
 QY 273 GTACAGATGGGAAGAGCTGAGTCCAGCTCAAAAGTGGCATGGAGGTCAAAGGTG 332
 DB 31738 GTACAGATGGGAAGAGCTGAGTCCAGCTCAAAAGTGGCATGGAGGTCAAAGGTG 31797
 QY 333 CACAGGAGGTGGTGTGCCACAGGGGGCAAGCAGATGCTGGGAGTACAGCTGTGA 392
 DB 31798 CACAGGAGGTGGTGTGCCACAGGGGGCAAGCAGATGCTGGGAGTACAGCTGTGA 31857
 QY 393 GGCTGGGGCCAGAGAGTCTCTTCCACCTGCACATCAG 433
 DB 31858 GGCTGGGGCCAGAGAGTCTCTTCCACCTGCACATCAG 31898

RESULT 5

AR542178

LOCUS

AR542178 2768 bp DNA linear PAT 08-OCT-2004

DEFINITION

Sequence 426 from patent US 6743619.

ACCESSION

AR542178

VERSION

AR542178.1 GI:53934258

KEYWORDS

Unknown.

SOURCE

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 2768)

AUTHORS

Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F.,

Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.J., Wehrman, T., Wang, J.-R.,

Wang, D. and Drmanac, R.T.

Nucleic acids and polypeptides

Patent: US 6743619-A 426 01-JUN-2004;

Location/Qualifiers

1..2768

/organism="unknown"

/mol_type="genomic DNA"

Best Local Similarity 57.4%; Pred. No. 4.4e-40;

Matches 510; Conservative 0; Mismatches 366; Indels 12; Gaps 3;

QY 75 GCACAGCAGCGCAAGACGGATGCGGGGACTACAGCTGCGAGGCCAGGGGCGCAGAGGT 134

DB 1874 GGAGGTGCGCCCGCAGCTGTCACGGGGCGCAGCTGAAGTTCGTGGCAACGGCATTGAGAG 1933

QY 135 CTCCTTCGGCTGCACATCACAGAGCCCAAGATGATGTTTGCNAAGGAGCAGTCAAGTCA 194

DB 1934 CAGCATCGGATGAGGTTCGGGGCGGCCAGGGCTGATCCCAACAAGCGCCACCGC 1993

QY 195 TAAAT-----GAGGTGAGGTGAGGGGGGGCGGCGAGTGCCTGCTGAGCTGTGAGGTGGC 248

DB 1994 AGTGGCGGGAGGTGCTGGCTCGCTGCACAGAGGCGGCGAGCTGCTGGCTGAGCTGTC 2053

QY 249 CCAGGCCCGACGAGGTGACGTGTACAAGGATGGGAAGAGCTGAGCTCCAGCTCAAA 308

DB 2054 AGATCAGGCTGCGGCTGTGACGTGGTGAAGGATGGTGCACACTGTCTCCAGGCCCAA 2113

QY 309 AGTGGGATGAGGTCAAGGGTGCACAGGAGCTGCTGTCACAGGCGGGGCAAGC 368

DB 2114 GTATGAGGTGAGGATCGGCCGGCGGGTGTCTCTTGTGCGAGATGTGGCCCGGA 2173

QY 369 AGATGCTGGGAGTACAGCTGTGAGGTGGGGGCCAGAGAGTCTCTTCCACCTGCACAT 428

DB 2174 CGATGAGGCTCTACGAGTGCCTCAGCCGGGGGCGCATCGCTACAGCTCTCCGT 2233

QY 429 CACAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATATAGGTGTCAGGCTGA 488

DB 2234 GCAAGGCTCGCGCTTCTTGCACAAGGACATGGCGGCGAGTGT---GTGGATGCGGT 2290

QY 489 GCGGGGACACTGCCATGCTGAGCTGTGAGTGGCCCGCCAGCCCGAGGAGTCAAGCTG 548

DB 2291 GGCTGGGGGCGCGCAGTGTGAGTGTGAGACCTCCGAAGCCCAAGCTCCACGTCACCTG 2350

QY 549 GTACAAGGAGCGGAAGAGTGCCTCCAGCTCAAAAGTACGATGAGGTCAAGGGCTG 608

DB 2351 GTACAAGATGGCATGGAGCTGGGCGCACTCCGTTGAGCGCTTCTTTCAGGAGGATGTGGG 2410

QY 609 CACAGC---AAGGCTGGTAGTGCAGAGGTGGGCAAGCAGATGCTGGGGAGTACAGCTG 665

DB 2411 GACGCGCACCGCTGGTGGCAGCCACAGTCAACAGGAGGATGAAGCACCCTACTCTG 2470

QY 566 CGAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGAGTGT 725

DB 2471 CCGCTGGGGGAGGACTCTGTGGACTTCGCGCTCCGCTCTCTGAGCCCAAGCGGTGT 2530

QY 726 TGCCAAGGAGCAGTTGTGTCATATGAGTGCAGGCTGAGGCGGCGCAGTGCACACT 785

DB 2531 TGCCAAGGAGCAGCGCGCTGACGGAGGTGACGCTGAGTGGGGCCAGCCCGCT 2590

QY 786 GAGCTGTGAGTGGCCCGCCAGGCCACAGAGAGGTGACGTGTACAAGGATGGGAAGAGCT 845

DB 2591 GAGCTGTGAGTGGCCCGCCAGGACCAGATGAGGTGACGTGTACAAGGAGGAGAGTT 2650

QY 846 GAGCTCAGTTCGAAAGTGCOCATAGAGGCTGCGGGCTGCATGCGGAGCTGTGTTGTCGA 905

DB 2651 GAGCTCCAGCTCGAAAGTGCACGCTGGAGGCTGTGGGCTGTATGCGGAGGCTGTGTTGTCGA 2710

QY 906 GCAGGAGCGCCAGGAGTGTGGGGAGTACACCTGTGAGGCTGGGG 953

DB 2711 GCAGGTGGCCAGGAGACTCCGATAGTACAGCTGTGAAGCCAGGTG 2758

RESULT 6

BD160445

LOCUS

BD160445

DEFINITION

Primer for synthesizing full-length cDNA and use thereof.

ACCESSION

BD160445

VERSION

BD160445.1

KEYWORDS

JP 2002191363-A/15288.

SOURCE

Homo sapiens (human)

ORGANISM

Query Match 24.8%; Score 248.4; DB 6; Length 2768;


```

/gene="OBSCN"
/number=7
2578..2859
/gene="OBSCN"
/number=8
2860..3169
/gene="OBSCN"
/number=8
3170..3445
/gene="OBSCN"
/number=9
3446..5480
/gene="OBSCN"
/number=9
5481..5750
/gene="OBSCN"
/number=10
5751..10615
/gene="OBSCN"
/number=10
10616..10891
/gene="OBSCN"
/number=11

```

ORIGIN

```

Query Match      21.8%; Score 218.6; DB 9; Length 24545;
Best Local Similarity 86.1%; Pred. No. 3.2e-34;
Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```

```

QY 705 CACAGGCCCGCCAGGAGCTGTTGCCAAGGAGGAGTGGTGCATATAGTGGCGACTGA 764
Db 10611 CCCAGAGCCCGCCAGGAGTGGTGGTAAAGGAGGAGCTGGCAGCGAGGAGCTGCAGGCCAGA 10670
QY 765 GGCAGGGCCGAGTGCACACTGAGCTGTGAGGTGGCCAGGCCAGCAGACAGAGGTGACGTG 824
Db 10671 GGCAGAGCCAGTGCACACTGAGCTGGAGGTGGCCAGGCCAGCAGCGAGGTGACGTG 10730
QY 825 GTACAAGATGGGAAGAAGTGTGAGCTTCAGTTCGAAAGTGCGCATAGAGGTCGCGGGCTG 884
Db 10731 GTACAAGATGGGAAGAAGTGTGAGCTTCAGTTCGAAAGTGTGCATGGAGGCCACAGGGCTG 10790
QY 885 CATCGGCAGCTGGTGTGCAGCAGGAGGCCAGCAGATCTGGGGAGTACACTGTGA 944
Db 10791 CACGGCAGGCTGGTGTGTGCAGCAGGAGGCCAGGCGATGCCGGGAGTATAGCTGCGA 10850
QY 945 GCCTGGGGGCCAGCGGCTCTCTCTCCACTCGATCTTTTCAG 985
Db 10851 GCCTGGGGGCCAGCGGCTCTCTCTCCACTCGATCTTTTCAG 10891

```

RESULT 10

```

AC023889
LOCUS Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
DEFINITION AC023889 174612 bp DNA linear HTG 07-JUL-2000
SEQUENCE, 13 unordered pieces.
ACCESSION AC023889
VERSION AC023889.3 GI:8969253
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 174612)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Genome Sequencing Center, Washington
AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 7, 2000 this sequence version replaced gi:8748947.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0661B12
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 14800; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 2478: contig of 2478 bp in length
* 2479: gap of unknown length
* 2579: contig of 4370 bp in length
* 7048: gap of unknown length
* 7049: contig of 2325 bp in length
* 9373: gap of unknown length
* 9374: contig of 3830 bp in length
* 9474: gap of unknown length
* 13303: contig of 3830 bp in length
* 13304: gap of unknown length
* 13404: contig of 4867 bp in length
* 18270: gap of unknown length
* 18371: contig of 7219 bp in length
* 25589: gap of unknown length
* 25590: contig of 9126 bp in length
* 25690: gap of unknown length
* 34815: contig of 10827 bp in length
* 34916: gap of unknown length
* 45742: contig of 17145 bp in length
* 45843: gap of unknown length
* 62987: contig of 17145 bp in length
* 62988: gap of unknown length
* 63088: contig of 24497 bp in length
* 75884: gap of unknown length
* 76884: contig of 27081 bp in length
* 87685: gap of unknown length
* 114766: contig of 27469 bp in length
* 114866: gap of unknown length
* 142335: contig of 32178 bp in length.
* 142435: gap of unknown length
* 142435: contig of 32178 bp in length.

FEATURES
Location/Qualifiers
1..174612
1..2478
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-661B12"
1..2478
/note="assembly_name:Contig4"
2579..6948
/note="assembly_name:Contig5"
7049..9373
/note="assembly_name:Contig6"
9474..13303
/note="assembly_name:Contig7"
13404..18270
/note="assembly_name:Contig8"
18371..25589
/note="assembly_name:Contig9"
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```

```

misc_feature 25690..34815
              /note="assembly_name:Contig10"
misc_feature 34916..45742
              /note="assembly_name:Contig11"
misc_feature 45843..62987
              /note="assembly_name:Contig12
clone end:SP6
vector_side:right"
misc_feature 63088..87584
              /note="assembly_name:Contig13"
misc_feature 87685..114765
              /note="assembly_name:Contig14"
misc_feature 114866..142334
              /note="assembly_name:Contig15"
misc_feature 142435..174612
              /note="assembly_name:Contig16"

ORIGIN
Query Match      21.8%; Score 218.6; DB 2; Length 174612;
Best Local Similarity 86.1%; Pred. No. 2.3e-34;
Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 705 CACAGAGCCCAAGGCAGTGTGTTGCCAAGGAGCAGTGTGTCATATGAGTGCGGACTGA 764
DB 95992 CCCAGAGCCCAAGGTGTGTTGCCAAGGAGCAGCAGCACACAGGAGGTGCAGGCTGA 96051
QY 765 GCGAGGGGCCAGTGCACACTAGCTGTGAGTGGCCAGGCCAGCAGAGGTCACGTG 824
DB 96052 GCGGGGGGCCAGTGCACGCTGAGCTGGAGGTGCGCCAGGCCAGCAGAGGTGACGCTG 96111
QY 825 GTACAAGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCAGTAGAGGTCGCGGCTG 884
DB 96112 GTACAAGATGGGAAGAGCTGAGTTCAGCTCGAAAGTGGCGTGGAGGCGGTGGGCTG 96171
QY 885 CATGGGCGAGCTGTGTGTCAGCAGCGCAGGCGCAGCAGATGTCGGGAGTACACCTGTGA 944
DB 96172 CACAGGAGGCTGTGTGTCAGCAGCGCGGCCAGCAGCGCGGGAGTACAGTGTGGA 96231
QY 945 GCGTGGGGCCAGCGGCTCTCTTCCACTGTGATTTTCAG 985
DB 96232 GCGAGGGGTGAGCAGCTCTCTTCCGCTGCGAGTGGCGAG 96272

RESULT 11
LOCUS CQ851220 3956 bp DNA linear PAT 23-AUG-2004
DEFINITION Sequence 1689 from Patent EP1447413.
ACCESSION CQ851220
VERSION CQ851220.1 GI:51509432
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isegai,T., Yamamoto,J., Nishikawa,T., Isono,Y., Sugiyama,T.,
Otsuki,T., Wakamatsu,A., Ishii,S., Nagai,K. and Irie,R.
TITLE Full-length human cDNA
JOURNAL Patent: EP 1447413-A 1689 18-AUG-2004;
Research Association for Biotechnology (JP)
FEATURES
source
Location/Qualifiers
1..3956
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match      21.8%; Score 218; DB 6; Length 3956;
Best Local Similarity 84.5%; Pred. NO. 5.7e-34;
Matches 245; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 144 CCTGCATCAGAGCCCAAGATCATCTTTCCAAAGGAGCAGTCAGTGCATATGAGGT 203

```

```

Db 1141 CTTGTCCATCCAGAGCCCAAGGTGGTGTTCGCCAAGGAGCAGCGCGCATGCAGGAGGT 1082
QY 204 CGAGGCTGAGCGGGGGCCACTGTCATGCTGAGCTGTGAGTGGGCCAGGCCAGACGGA 263
Db 1081 GCAGCGGAGCGGGGGCCAGTGCACGCTGAGCTGTGAGTGGGCCAGGCGCAGATGA 1022
QY 264 GGTGACGTGGTACAGGATGGAGAGAGCTCAGCTCCAGCTCAAAGTGGGCATGGAGGT 323
Db 1021 GGTGACATGGTACAGGAGCGGGAAGAGCTGAGCTCCAGCTCAAAGTGGCGATGGAGG 962
QY 324 CAAAGGGTGCACACGAGGCTGTGCTGCCACAGCGGGGCAAGCAGATGCTGGGGAGTA 383
Db 961 CAGCGGCTACACACGAGGCTGTGAGTGCACAGCGGGGCCAGGCGGATGCTGGGGAGTA 902
QY 384 CAGCTGTGAGCTGGGGGCCAGAGAGTCTCTTCCACTGCACATCACAG 433
Db 901 CAGCTGTGAGCGGGGGCCCAACGCTGTCTTCCGCTGCGACGTCGCGAG 852

RESULT 12
LOCUS AK128447/c 3956 bp mRNA linear PRI 19-FEB-2004
DEFINITION Homo sapiens cDNA FLJ46590 fis, clone THYMU304441.
ACCESSION AK128447
VERSION AK128447.1 GI:34535823
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
Kodaira,H., Furiya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A.,
Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K.,
Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y.,
Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isegai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3956)
AUTHORS Isegai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2003) Takao Isegai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: Reverse Proteomics Research Institute, HRI and
RAB.
FEATURES
source
Location/Qualifiers
1..3956
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYMU304441"
/tissue type="thymus"
/clone_lib="THYMU3"
/note="Cloning vector: pME18SFL3"
1014..1415
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC87445.1"
/db_xref="GI:34535824"
/translacion="MSPSPGPPPHSSAHWPPPPPPAPPAPAPAPWOTPPNALGHTRR
DAGSEAPEQTQPMQCPSQSDSGLSIKQCTVTHRAPPTTRAKGLRLHVGSTIIMSLG

```


DEFINITION Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT
SEQUENCE, 31 unordered pieces.

ACCESSION AC026657
VERSION AC026657.4 GI:9958202
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS Waterston,R.H.
REFERENCE 2 (bases 1 to 164766)
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT On Sep 1, 2000 this sequence version replaced gi:7637349.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0245P10
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: Plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 139884 bases at least Q40
Consensus quality: 147686 bases at least Q30
Consensus quality: 151469 bases at least Q20
Insert size: 169000; agarose-fp
Insert size: 161074; sum-of-contigs
Quality coverage: 3.60 in Q20 bases; agarose-fp
Quality coverage: 3.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1124: contig of 1124 bp in length
* 1125 1224: gap of unknown length
* 1225 3032: contig of 1808 bp in length
* 3032 3033: gap of unknown length
* 3033 3132: contig of 1361 bp in length
* 3132 4494: gap of unknown length
* 4494 4594: gap of unknown length
* 4594 5861: gap of 1267 bp in length
* 5861 5960: gap of unknown length
* 5960 7671: contig of 1710 bp in length
* 7671 7771: gap of unknown length
* 7771 9670: contig of 1899 bp in length
* 9670 9769: gap of unknown length
* 9769 11715: contig of 1946 bp in length
* 11715 11816: gap of unknown length
* 11816 14243: contig of 2428 bp in length
* 14243 14344: gap of unknown length
* 14344 16888: contig of 2544 bp in length
* 16888 16987: gap of unknown length
* 16987 19247: contig of 2260 bp in length
* 19247 19348: gap of unknown length
* 19348 21375: contig of 2028 bp in length
* 21375 21475: gap of unknown length
* 21475 25025: contig of 3550 bp in length
* 25025 25125: gap of unknown length
* 25125 25026

```
* 25126 28051: contig of 2926 bp in length
* 28052 28151: gap of unknown length
* 28152 32054: contig of 3903 bp in length
* 32055 32154: gap of unknown length
* 32155 35716: contig of 3562 bp in length
* 35717 35816: gap of unknown length
* 35817 40891: contig of 5075 bp in length
* 40892 40991: gap of unknown length
* 40992 44027: contig of 3036 bp in length
* 44028 44127: gap of unknown length
* 44128 48899: contig of 4772 bp in length
* 48900 54255: gap of unknown length
* 54256 54355: gap of unknown length
* 54356 58994: contig of 4639 bp in length
* 58995 59094: gap of unknown length
* 59095 64390: contig of 5296 bp in length
* 64391 70865: contig of 6375 bp in length
* 70866 78667: gap of unknown length
* 78668 87207: contig of 8440 bp in length
* 87208 87308: gap of unknown length
* 87309 97959: gap of unknown length
* 97960 109520: contig of 11562 bp in length
* 109521 120908: gap of unknown length
* 120909 121009: contig of 11288 bp in length
* 121010 141477: contig of 20469 bp in length
* 141478 162672: contig of 21095 bp in length
* 162673 162773: gap of unknown length
* 162774 163975: contig of 1202 bp in length
* 163976 164075: gap of unknown length
* 164076 164766: contig of 692 bp in length.
FEATURES
    source
        1..164766
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-245P10"
        1..1124
            /note="assembly_name:Contig14"
        1225..3032
            /note="assembly_name:Contig21"
        3133..4493
            /note="assembly_name:Contig23"
        4594..5860
            /note="assembly_name:Contig24"
        5961..7670
            /note="assembly_name:Contig25"
        7771..9669
            /note="assembly_name:Contig26"
        9770..11715
            /note="assembly_name:Contig27"
        11816..14243
            /note="assembly_name:Contig28"
        14344..16887
            /note="assembly_name:Contig29"
        16988..19247
            /note="assembly_name:Contig30"
        19348..21375
            /note="assembly_name:Contig31"
        21476..25025
            /note="assembly_name:Contig32"
        25126..28051
            /note="assembly_name:Contig33"
        28152..32054
            /note="assembly_name:Contig34"
        32155..35716
```

```
misc_feature /note="assembly_name:Contig35" 35817..40891
misc_feature /note="assembly_name:Contig36" 40992..44027
misc_feature /note="assembly_name:Contig37" 44128..48899
misc_feature /note="assembly_name:Contig38" 49000..54255
misc_feature /note="assembly_name:Contig39" 54356..58994
misc_feature /note="assembly_name:Contig40" 59095..64390
misc_feature /note="assembly_name:Contig41" 64491..70865
misc_feature /note="assembly_name:Contig42" 70966..78667
misc_feature /note="assembly_name:Contig43" 78768..87207
misc_feature /note="assembly_name:Contig44" 87308..97858
misc_feature /note="assembly_name:Contig45" 97959..109520
misc_feature /note="assembly_name:Contig46" 109621..120908
misc_feature /note="assembly_name:Contig47" 121009..141477
misc_feature /note="assembly_name:Contig48" 141578..162672
misc_feature /note="assembly_name:Contig49" 162773..163974
misc_feature /note="assembly_name:Contig9" 164075..164766
misc_feature /note="assembly_name:Contig7"

ORIGIN
Query Match      20.7%; Score 207.6; DB 2; Length 164766;
Best Local Similarity 84.5%; Pred. No. 3.9e-32;
Matches 245; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

Qy 144 CTTGCATCACAGAGCCCAAGATGATGTTTGCNAAGGAGCAGTCAGTGCATATAGGT 203
Db 56107 CTTGTCCATCCAGAGCCCAAGGTGGTGTTCACAGGAGCAGCCGCGCATGCGAGGT 56166

Qy 204 GCAGGCTGAGCGGGGCCAGTGCCATGCTGAGCTGTGAGTGGCCAGCCAGACCGGA 263
Db 56167 GCAGGCGAGCGGGGCCAGTGCCAGCTGAGCTGTGAGTGGCCAGCCAGATGGA 56226

Qy 264 GTTGACGTGTGTACAGGATGGGAAGAGCTGAGCTCCAGTCAAAAGTGGGCATGGAGGT 323
Db 56227 GGTGACATGGTACAAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGGC 56286

Qy 324 CAAAGGCTGCACACGAGGCTGGTGTCCACAGGGGGCAAGAGCAGATGCTGGGAGTA 383
Db 56287 CAGCGGCTACACACGAGGCTGGTGTGAGCAGAGGGGGCCAGCGGATGCTGGGAGTA 56346

Qy 384 CAGCTGTAGGCTGGGGGCCAGAGAGTCTCTTCCACTGCATCACAG 433
Db 56347 CAGCTGTAGGCTGGGGGCCAAGCGGTGTCTTCCGCTGCACGTGGCAG 56395
```

Search completed: March 21, 2005, 05:34:31
Job time : 4521.62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:32:32 ; Search time 627.052 Seconds
(without alignments)
9503.485 Million cell updates/sec

Title: US-10-077-130-4_COPY_22500_23500

Perfect score: 1001
Sequence: 1 ggcacgtcagtgacactggc.....ctgtcgaagtcgtggacct 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5544816 seqs, 2976611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	3225	17	US-10-311-034-46
2	1001	100.0	4175	15	US-10-307-019-2
3	1001	100.0	4936	17	US-10-182-243-14
4	1001	100.0	5454	17	US-10-415-011-44
5	1001	100.0	7893	13	US-10-077-130-3
6	1001	100.0	8106	13	US-10-077-130-1
7	1001	100.0	23907	13	US-10-077-130-6
8	1001	100.0	24120	13	US-10-077-130-6
9	999.4	99.8	5007	15	US-10-307-019-3
10	999.4	99.8	5207	9	US-09-858-664A-1
11	999.4	99.8	5207	14	US-10-274-978-1

12	999.4	99.8	5207	14	US-10-274-978-3
13	999.4	99.8	5207	17	US-10-697-263-1
14	999.4	99.8	5207	17	US-10-697-263-1
15	999.4	99.8	5207	18	US-10-921-168-1
16	999.4	99.8	5207	18	US-10-921-168-3
17	999.4	99.8	7928	15	US-10-307-019-5
18	696	69.5	1253	17	US-10-425-114-26278
19	151.6	15.1	9930	17	US-10-210-130-33
20	148.4	14.8	1627	17	US-10-425-114-26285
21	148.4	14.8	7789	17	US-10-333-314-38
22	146.8	14.7	9698	17	US-10-210-130-37
23	146.8	14.7	10122	17	US-10-210-130-35
24	140.4	14.0	9807	17	US-10-379-381-1
c 25					13.1
c 26					13.1
c 27					13.1
c 28					13.0
c 29					13.0
c 30					10.5
c 31					9.8
c 32					9.2
c 33					9.2
c 34					9.0
c 35					9.0
c 36					8.6
c 37					8.6
c 38					8.5
c 39					8.5
c 40					8.5
c 41					8.4
c 42					8.4
c 43					8.3
c 44					8.3
c 45					8.3

ALIGNMENTS

RESULT 1
US-10-311-034-46
; Sequence 46, Application US/10311034
; Publication No. US20040023242A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: BANDMAN, Olga
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Yan
; APPLICANT: GANDHI, Aneena R.
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: YAO, Monique G.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: GREENWALD, Sara R.
; APPLICANT: RAMKUMAR, Javalaxmi
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: KEARNEY, Liam
; APPLICANT: BURFORD, Neil
; APPLICANT: NGUYEN, Danniell B.
; APPLICANT: TANG, Y. Tom
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HE, Ann
; APPLICANT: THORNTON, Michael
; APPLICANT: HAFALIA, April
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: LO, Terence P.
; APPLICANT: KHAH, Farrah A.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: AZIMZAI, Yalda

Query Match	100.0.0%;	Score 1001;	DB 17;	Length 3225;
Best Local Similarity	100.0.0%;	Pred. No. 2.e-261;		
Matches 1001;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GGCCAGTCAGTGAACACTGGCTCTCAGAGTGTCTAGCCAGCCAGCAGCTGCCAGGCCACCTGG	60	
DB	1663	GGCCAGTCAGTGAACACTGGCTCTCAGAGTGTCTAGCCAGCCAGCAGCTGCCAGGCCACCTGG	1722	
QY	61	AGCAAGAAGCGAGCCGCCCTTGGAGAGCAGAGCCGTGTCTCATCTCTCTGCCACCTTCAAG	120	
DB	1723	AGCAAGAAGCGAGCCGCCCTTGGAGAGCAGAGCCGTGTCTCATCTCTCTGCCACCTTCAAG	1782	
QY	121	AACCTCCAGCTTCTGACCACTCTCTGGTGTGGCTTGAGGACCTGGGTGTGTACACCTGC	180	
DB	1783	AACCTCCAGCTTCTGACCACTCTCTGGTGTGGCTTGAGGACCTGGGTGTGTACACCTGC	1842	
QY	181	AGCGTAGCAATCGCTGGGGACAGTGAACACCAACCGGGCTCTCCGGAAGGCAGAGCG	240	
DB	1843	AGCGTAGCAATCGCGCTGGGACAGTGAACACCAACCGGGCTCTCCGGAAGGCAGAGCG	1902	
QY	241	CCCTCATCTTCGCCATCGCCGGATATCGGGAGGTGTACCGGATGGGTGTCTGTGTCT	300	
DB	1903	CCCTCATCTTCGCCATCGCCGGATATCGGGAGGTGTACCGGATGGGTGTCTGTGTCT	1962	
QY	301	TGGAAGCCCGTGGAAATCTACTAGGCCCTGTGACCTCATTTGTGAGTCAGACCTTAGAAGGC	360	
DB	1963	TGGAAGCCCGTGGAAATCTACTAGGCCCTGTGACCTCATTTGTGAGTCAGACCTTAGAAGGC	2022	
QY	361	GGCAGCTGACACACACTGGCTCTCGACATCTTTGACTGTCTGTACTGTACACGACGAGCTC	420	
DB	2023	GGCAGCTGACACACACTGGCTCTCGACATCTTTGACTGTCTGTACTGTACACGACGAGCTC	2082	
QY	421	TCCCGGGGTGGCACTTACACCTTTCGACGCGATGTGACAAAGGCAGGAAATGGTCCC	480	
DB	2083	TCCCGGGGTGGCACTTACACCTTTCGACGCGATGTGACAAAGGCAGGAAATGGTCCC	2142	
QY	481	TACAGAGCCCTCTGGAGCAAGTCTCTCTGGAGGGCCAGCCACTGGCCCTCTGAGGAG	540	
DB	2143	TACAGAGCCCTCTGGAGCAAGTCTCTCTGGAGGGCCAGCCACTGGCCCTCTGAGGAG	2202	
QY	541	GAGAGCCAGGGCGGTACGCCAAACCCCTGCCAGCACAAAGACCTTGCATTTCCAGACA	600	
DB	2203	GAGAGCCAGGGCGGTACGCCAAACCCCTGCCAGCACAAAGACCTTGCATTTCCAGACA	2262	
QY	601	CAGATCCAGAGGGCGCTTTCAGCGTGTGTGCGCAATGTGGGAGAAAGCCAGCGGGCGG	660	
DB	2263	CAGATCCAGAGGGCGCTTTCAGCGTGTGTGCGCAATGTGGGAGAAAGCCAGCGGGCGG	2322	

Qy	241	CCCTCATCTTCGCCATGCCCGGATATACGGGAGGTGTACCGGATGGGGTCTGCTGGTC	300
Db	2818	CCCTCATCTTCGCCATGCCCGGATATACGGGAGGTGTACCGGATGGGGTCTGCTGGTC	2877
Qy	301	TGGAAGCCCGTGGAAATCCTACGGGCCCTGTGACCTTACATTTGTGCAGTGCAGCCTAGAGGC	360
Db	2878	TGGAAGCCCGTGGAAATCCTACGGGCCCTGTGACCTTACATTTGTGCAGTGCAGCCTAGAGGC	2937
Qy	361	GGCAGCTGGACCAACACTGGCCCTCCGACATCTTTGACTGTCTACCTTGACCCAGCAAGCTC	420
Db	2938	GGCAGCTGGACCAACACTGGCCCTCCGACATCTTTGACTGTCTACCTTGACCCAGCAAGCTC	2997
Qy	421	TCCCGGGGTGGCACTTACACTTCCGCACGGCATGTGTCAACAAAGGCAGAAATGGGTCCC	480
Db	2998	TCCCGGGGTGGCACTTACACTTCCGCACGGCATGTGTCAACAAAGGCAGAAATGGGTCCC	3057
Qy	481	TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGCCACGCCACCTTGGCATTTCAGAGG	540
Db	3058	TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGCCACGCCACCTTGGCATTTCAGAGG	3117
Qy	541	GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCCCAGCACAAGACCTTTGCGATTTCCAGACA	600
Db	3118	GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCCCAGCACAAGACCTTTGCGATTTCCAGACA	3177
Qy	601	CAGATCAGAGGGGCCGCTTACGGTGGTGGGCAATGCTGGAGAAAGGCAGCGGGCGG	660
Db	3178	CAGATCAGAGGGGCCGCTTACGGTGGTGGGCAATGCTGGAGAAAGGCAGCGGGCGG	3237
Qy	661	GCCTGGCCGCGCAGATCATCCCTACCCACCCCAAGCACAAGACAGCAGTGTGCGCGAA	720
Db	3238	GCCTGGCCGCGCAGATCATCCCTACCCACCCCAAGCACAAGACAGCAGTGTGCGCGAA	3297
Qy	721	TACGAGGCCCTCAAGGGCCTGCGCACCCCGCACCTGGCCCGCAGTGTGACGAGCCTACCTC	780
Db	3298	TACGAGGCCCTCAAGGGCCTGCGCACCCCGCACCTGGCCCGCAGTGTGACGAGCCTACCTC	3357
Qy	781	AGCCCCGGCACCCTGGTGTCTATTGGAGTGTGTCTCTGGGCCGAGCTGTCTCCCTGC	840
Db	3358	AGCCCCGGCACCCTGGTGTCTATTGGAGTGTGTCTCTGGGCCGAGCTGTCTCCCTGC	3417
Qy	841	CTGGCCGAGAGGGCCTCTACTCAGATCCGAGGTGAAGACTACCTGTGGCAGATGTTG	900
Db	3418	CTGGCCGAGAGGGCCTCTACTCAGATCCGAGGTGAAGACTACCTGTGGCAGATGTTG	3477
Qy	901	AGTGCCACCCAGTACCTGCAACACGACATCTCTGCACCTGGACCTTGAGTCCGAGAAC	960
Db	3478	AGTGCCACCCAGTACCTGCAACACGACATCTCTGCACCTGGACCTTGAGTCCGAGAAC	3537
Qy	961	ATGATCATACCGAATCAACCTGCTCAAGGTGTGGACT	1001
Db	3538	ATGATCATACCGAATCAACCTGCTCAAGGTGTGGACT	3578

RESULT 3

```

US-10-182-243-14
; Sequence 14, Application US/10182243
; Publication No. US20040048310A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY D.
; APPLICANT: WHYTE, DAVID
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; APPLICANT: MARTINEZ, RICARDO
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE
; TITLE OF INVENTION: ENZYMES
; FILE REFERENCE: 038602/1366
; CURRENT APPLICATION NUMBER: US/10/182,243
; CURRENT FILING DATE: 2003-07-07
; PRIOR APPLICATION NUMBER: PCT/US01/02337
; PRIOR FILING DATE: 2001-01-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14

```

Qy 961 ATGATCATCCGGAATACAACTGCTCAAGTCTGTGGACCT 1001
Db ||||||||||||||||||||||||||||||||||||||||||||
4339 ATGATCATCCGGAATACAACTGCTCAAGTCTGTGGACCT 4379

RESULT 4

US-10-415-011-44
; Sequence 44, Application US/10415011
; Publication No. US20040053394A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: DING, Li
; APPLICANT: TANG, Y. Tom
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: NGUYEN, Dannie B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LU, Yan
; APPLICANT: YUE, Henry
; APPLICANT: BURFORD, Neil
; APPLICANT: BANDMAN, Olga
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: LAL, Preeti G.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: KHAN, Farrah A.
; APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0262 USN
; CURRENT APPLICATION NUMBER: US/10/415,011
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/47728
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,410
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/244,068
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/245,708
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/247,672
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/249,565
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,730
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/250,807
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PERL Program
; SEQ ID NO 44
; LENGTH: 5454
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040053394A1 7638121CB1
US-10-415-011-44

Query Match 100.0%; Score 1001; DB 17; Length 5454;
Best Local Similarity 100.0%; Pred. No. 2.3e-261;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCAGTCACTGACACTGGCCTGCCAGGTGTACGCCAGCCAGCTGCCAGGCCACCTGG 60
Db ||||||||||||||||||||||||||||||||||||||||||||
3892 GGCAGTCACTGACACTGGCCTGCCAGGTGTACGCCAGCCAGCTGCCAGGCCACCTGG 3951
Qy 61 AGCAAGACGGAGCCCCCTGGAGAGCAGCAGCGGTCTCTCATCTCTCCACCCCTCAAG 120
Db ||||||||||||||||||||||||||||||||||||||||||||
3952 AGCAAGACGGAGCCCCCTGGAGAGCAGCAGCGGTCTCTCATCTCTGCCACCCTCAAG 4011
Qy 121 AACTCCAGCTTCTGACCATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 180
Db ||||||||||||||||||||||||||||||||||||||||||||
4012 AACTCCAGCTTCTGACCATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4071
Qy 181 AGCGTGAGCAATCGCTGGGGACAGTACCAACACCGGGGTCTCTCCGAGAGCAGAGCGC 240
Db ||||||||||||||||||||||||||||||||||||||||||||
4072 AGCGTGAGCAATCGCTGGGGACAGTACCAACACCGGGGTCTCTCCGAGAGCAGAGCGC 4131
Qy 241 CCCTCATCTTCGCATGCCCGGATATCGGGAGGTGTACCGGATGGGGTCTGCTGCTC 300
Db ||||||||||||||||||||||||||||||||||||||||||||
4132 CCCTCATCTTCGCATGCCCGGATATCGGGAGGTGTACCGGATGGGGTCTGCTGCTC 4191
Qy 301 TGGAGCCCGTGGAAATCTTACGGCCCTGTACCTTACATTTGTGCAGTGCAGCTTAGAAGGC 360
Db ||||||||||||||||||||||||||||||||||||||||||||
4192 TGGAGCCCGTGGAAATCTTACGGCCCTGTACCTTACATTTGTGCAGTGCAGCTTAGAAGGC 4251
Qy 361 GGCAGCTGGACCAACACTGGGCTTCGACATCTTTTGTGCTGTCTTACCTGACAGCAAGCTC 420
Db ||||||||||||||||||||||||||||||||||||||||||||
4252 GGCAGCTGGACCAACACTGGGCTTCGACATCTTTTGTGCTGTCTTACCTGACAGCAAGCTC 4311
Qy 421 TCCGGGGTGGACCTACACTTCCGCACGGCATGTGTACAGAGGAGGAGGAGGAGGAGGAGG 480
Db ||||||||||||||||||||||||||||||||||||||||||||
4312 TCCGGGGTGGACCTACACTTCCGCACGGCATGTGTACAGAGGAGGAGGAGGAGGAGGAGG 4371
Qy 481 TACAGACAGCCCTCGGAGCAAGTCTCTCGGAGGGCCAGCCACCTGGGCTCTTGAGGAG 540
Db ||||||||||||||||||||||||||||||||||||||||||||
4372 TACAGACAGCCCTCGGAGCAAGTCTCTCGGAGGGCCAGCCACCTGGGCTCTTGAGGAG 4431
Qy 541 GAGAGCAGGGGGGGTTCAGGCCAACCTCCGCCAGCAAGAGCTTCGCAATTCAGACACA 600
Db ||||||||||||||||||||||||||||||||||||||||||||
4432 GAGAGCAGGGGGGGTTCAGGCCAACCTCCGCCAGCAAGAGCTTCGCAATTCAGACACA 4491
Qy 601 CAGATCCAGAGGGGGCGCTTCAGCGTGTGGGCAATGTCTGGGAGAGAGGCGGGCGG 660
Db ||||||||||||||||||||||||||||||||||||||||||||
4492 CAGATCCAGAGGGGGCGCTTCAGCGTGTGGGCAATGTCTGGGAGAGAGGCGGGCGG 4551
Qy 661 GCGCTGGCCGCAAGATCATCCCTACACCCCAAGGACAAAGACAGCAGTCTGCGCGAA 720
Db ||||||||||||||||||||||||||||||||||||||||||||
4552 GCGCTGGCCGCAAGATCATCCCTACACCCCAAGGACAAAGACAGCAGTCTGCGCGAA 4611
Qy 721 TACGAGCCCTCAAGGGCTGGCCACCGCACCTGGCCAGCTGCACGAGCCCTACCTC 780
Db ||||||||||||||||||||||||||||||||||||||||||||
4612 TACGAGCCCTCAAGGGCTGGCCACCGCACCTGGCCAGCTGCACGAGCCCTACCTC 4671
Qy 781 AGCCCCGGCACCTGGTGTCTCATCTTGAGCTGTGTCTCTGGGCGCGAGCTGTCCCTTCG 840
Db ||||||||||||||||||||||||||||||||||||||||||||
4672 AGCCCCGGCACCTGGTGTCTCATCTTGAGCTGTGTCTCTGGGCGCGAGCTGTCCCTTCG 4731
Qy 841 CTGCGCGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTGCGAGATGTTG 900
Db ||||||||||||||||||||||||||||||||||||||||||||
4732 CTGCGCGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTGCGAGATGTTG 4791
Qy 901 AGTGCCACCCAGTACCTGCACAAACAGCACATCTTGCACTGGAGCTGAGGTCCGAGAAC 960
Db ||||||||||||||||||||||||||||||||||||||||||||
4792 AGTGCCACCCAGTACCTGCACAAACAGCACATCTTGCACTGGAGCTGAGGTCCGAGAAC 4851
Qy 961 ATGATCATCACCGGAATACAACTCTGCTCAAGGTCTGGACCT 1001
Db ||||||||||||||||||||||||||||||||||||||||||||
4852 ATGATCATCACCGGAATACAACTCTGCTCAAGGTCTGGACCT 4892

RESULT 5
US-10-077-130-3
; Sequence 3, Application US/10077130
; Publication No. US20020168742A1

6786 TGAAGCCCGTGGAACTCTACGGCCCTGTGACCTACATTGTGAGTGCAGCCTAGAGGC 6845
Db
361 GGCAGCTGGACCAACACTGGCCCTCCGACATCTTTGACTGCTGTACTGTGACAGCAAGCTC 420
Qy
6846 GGCAGCTGGACCAACACTGGCCCTCCGACATCTTTGACTGCTGTACTGTGACAGCAAGCTC 6905
Db
421 TCCGGGGTGGCACTTACACCTTCGGACCGCATGTGTGACAGGAGGAGGAGTGGGTCCC 480
Qy
6906 TCCGGGGTGGCACTTACACCTTCGGACCGCATGTGTGACAGGAGGAGGAGTGGGTCCC 6965
Db
481 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Qy
6966 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7025
Db
541 GAGAGCCAGGGGCGGTGACGCCCAACCCCTGCCCAGCAAAAGACTTTCGCATTTCCAGACA 600
Qy
7026 GAGAGCCAGGGGCGGTGACGCCCAACCCCTGCCCAGCAAAAGACTTTCGCATTTCCAGACA 7085
Db
601 CAGATCCAGAGGGGCGGTTCAGCGTGTGGGCAATGCTGGGAGAGGCGGAGGCGG 660
Qy
7086 CAGATCCAGAGGGGCGGTTCAGCGTGTGGGCAATGCTGGGAGAGGCGGAGGCGG 7145
Db
661 GCGCTGGCGCCCAAGATCATCCCTTACCAACCCCAAGGACAAAGACAGCAGTGTCTGCGGAA 720
Qy
7146 GCGCTGGCGCCCAAGATCATCCCTTACCAACCCCAAGGACAAAGACAGCAGTGTCTGCGGAA 7205
Db
721 TACAGAGCCCTCAAGGGCGGTGGCCACCCGACCTGGCCAGCTGCACGAGCCTTACCTC 780
Qy
7206 TACAGAGCCCTCAAGGGCGGTGGCCACCCGACCTGGCCAGCTGCACGAGCCTTACCTC 7265
Db
781 AGCCCGCGCACCTGGTCTCATCTTGGAGCTGTCTCTGGCCCGAGCTGTCTCCCTGC 840
Qy
7266 AGCCCGCGCACCTGGTCTCATCTTGGAGCTGTCTCTGGCCCGAGCTGTCTCCCTGC 7325
Db
841 CTGGCGCGAGAGGCGCTCTTACTCAGAACTCCGAGGTGAAGGACTACTGTGGCAGATGTTG 900
Qy
7326 CTGGCGCGAGAGGCGCTCTTACTCAGAACTCCGAGGTGAAGGACTACTGTGGCAGATGTTG 7385
Db
901 AGTGCCACCCAGTACTGTGACAAACAGCAGACATCTCGACCTGGACCTGAGGTCCGAGAAC 960
Qy
7386 AGTGCCACCCAGTACTGTGACAAACAGCAGACATCTCGACCTGGACCTGAGGTCCGAGAAC 7445
Db
961 ATGATCATCAGCAATACAACTGTCTCAAGGTCTGGACCT 1001
Qy
7446 ATGATCATCAGCAATACAACTGTCTCAAGGTCTGGACCT 7486
Db

RESULT 7

US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(W)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match 100.0%; Score 1001; DB 13; Length 23907;
Best Local Similarity 100.0%; Pred. No. 2,7e-261;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8

US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1

1 GGCAGTCACTGACACTGGCCCTGCCAGTGTTCAGGCCAGCAGTGTGCCAGGCCACCTGG 60
Qy
22429 GGCAGTCACTGACACTGGCCCTGCCAGTGTTCAGGCCAGCAGTGTGCCAGGCCACCTGG 22488
Db
61 AGCAAAAGACGAGCCCCCTTGGAGAGCAGCGGTCTCTCATCTCTGCCACCTCAAG 120
Qy
22489 AGCAAAAGACGAGCCCCCTTGGAGAGCAGCGGTGTCTCATCTCTGCCACCTCAAG 22548
Db
121 AACTTCCAGTCTTGCACCATCTCTGGTGTGTGGCTGAGGACCTGGGTGTACACCTGC 180
Qy
22549 AACTTCCAGTCTTGCACCATCTCTGGTGTGTGGCTGAGGACCTGGGTGTACACCTGC 22608
Db
181 AGCGTGAGCAATGCGTGGGGACAGTGACCAACACCGGGGTCTCTCGGAGGAGAGCGC 240
Qy
22609 AGCGTGAGCAATGCGTGGGGACAGTGACCAACACCGGGGTCTCTCGGAGGAGAGCGC 22668
Db
241 CCTCATCTTCGCGCATGCGCGGATATCGGGAGGTGTACGCGATGGGGTGTCTGGTGC 300
Qy
22669 CCTCATCTTCGCGCATGCGCGGATATCGGGAGGTGTACGCGATGGGGTGTCTGGTGC 22728
Db
301 TGAAGCCCGTGGAACTCTACGGCCCTGTGACCTACATTGTGAGTGCAGCCTTAGAGGC 360
Qy
22729 TGAAGCCCGTGGAACTCTACGGCCCTGTGACCTACATTGTGAGTGCAGCCTTAGAGGC 22788
Db
361 GGAGCTGGACCAACACTGGGCTCCGACATCTTTGACTGTCTGCTGACCTGACCAAGCTC 420
Qy
22789 GGAGCTGGACCAACACTGGGCTCCGACATCTTTGACTGTCTGCTGACCTGACCAAGCTC 22848
Db
421 TCCGGGGTGGCACTTACACTTCCGACCGCATGTGTGAGGAGGAGGAGGAGGAGGAGGAG 480
Qy
22849 TCCGGGGTGGCACTTACACTTCCGACCGCATGTGTGAGGAGGAGGAGGAGGAGGAGGAG 22908
Db
481 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Qy
22909 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 22968
Db
541 GAGAGCAGGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACTTTCGATTTCCAGACA 600
Qy
22969 GAGAGCAGGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACTTTCGATTTCCAGACA 23028
Db
601 CAGATCCAGAGGGGCGGTGAGCGTGTGGCAATGTCTGGGAGAGGAGGAGGAGGAGGAGGAG 660
Qy
23029 CAGATCCAGAGGGGCGGTGAGCGTGTGGCAATGTCTGGGAGAGGAGGAGGAGGAGGAGGAG 23088
Db
661 GCGCTGGCGCAAGATCATCTCTTACCAACCCAGGAGCAGCAGCAGTGTCTGCGGAA 720
Qy
23089 GCGCTGGCGCAAGATCATCTCTTACCAACCCAGGAGCAGCAGCAGTGTCTGCGGAA 23148
Db
721 TACGAGGCCCTCAAGGGCGTGGCCACCCGACCTGGCCAGCTGCAGCAGCCTTACCTC 780
Qy
23149 TACGAGGCCCTCAAGGGCGTGGCCACCCGACCTGGCCAGCTGCAGCAGCCTTACCTC 23208
Db
781 AGCCCCCGGACCTGGTGTCTCATTTGAGTGTGTCTGGGCCCGAGCTGTCTCCCTGC 840
Qy
23209 AGCCCCCGGACCTGGTGTCTCATTTGAGTGTGTCTGGGCCCGAGCTGTCTCCCTGC 23268
Db
841 CTGGCCGAGAGGGCGCTCTTACTCAGAACTCCGAGGTGAAGGACTACTCTGGCAGATGTTG 900
Qy
23269 CTGGCCGAGAGGGCGCTCTTACTCAGAACTCCGAGGTGAAGGACTACTCTGGCAGATGTTG 23328
Db
901 AGTGCCACCCAGTACTGTGACCAACAGCAGACATCTCTGACCTGGACCTGAGGTCCGAGAAC 960
Qy
23329 AGTGCCACCCAGTACTGTGACCAACAGCAGACATCTCTGACCTGGACCTGAGGTCCGAGAAC 23388
Db
961 ATGATCATCAGCAATACAACTGTCTCAAGGTCTGGACCT 1001
Qy
23389 ATGATCATCAGCAATACAACTGTCTCAAGGTCTGGACCT 23429
Db

Qy	661	GGCTGGCCGCAAGATCATCCCTTACCACCCCAAGGACAAGACAGCAGTCTCTGCGAA	720
Db	23160	GGCTGGCCGCAAGATCATCCCTTACCACCCCAAGGACAAGACAGCAGTCTGCGCA	23219
Qy	721	TACGAGGCCCTCAAGGGCCTGCGCACCCGCACTTGGCCCAAGCTGCACGAGCCTACCTC	780
Db	23220	TACGAGGCCCTCAAGGGCCTGCGCACCCGCACTTGGCCCAAGCTGCACGAGCCTACCTC	23279
Qy	781	AGCCCCCGACACTGGTGTCTCATCTTGGAGCTGTCTCTGGCCCCGAGCTGCTCCCCCTGC	840
Db	23280	AGCCCCCGACACTGGTGTCTCATCTTGGAGCTGTCTCTGGCCCCGAGCTGCTCCCCCTGC	23339
Qy	841	CTGGCCGAGAGGGCCTCTTACTCAGAACTCCGAGGTGAAGGACTACTCTGGCAGATGTTG	900
Db	23340	CTGGCCGAGAGGGCCTCTTACTCAGAACTCCGAGGTGAAGGACTACTCTGGCAGATGTTG	23399
Qy	901	AGTGCCACCAGTACTCTGCACAACCAAGCACATCTTGCACTTGGACCTGAGTCCGAGAAC	960
Db	23400	AGTGCCACCAGTACTCTGCACAACCAAGCACATCTTGCACTTGGACCTGAGTCCGAGAAC	23459
Qy	961	ATGATCATCAGGAATACAACCTGTCAAGGTCGTGGACCT	1001
Db	23460	ATGATCATCAGGAATACAACCTGTCAAGGTCGTGGACCT	23500

```

RESULT 9
US-10-307-019-3
; Sequence 3, Application US/10307019
; Publication No. US2003010853A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; FILE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DV1
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5007
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (97)...(4926)
US-10-307-019-3

Query Match          99.8%; Score 999.4; DB 15; Length 5007;
Best Local Similarity 99.9%; Pred. No. 6.3e-261;
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1   GGCCAGTCAGTGCACACTGGCCCTGCCAGGTGTCAGCCAGCAGCGTGCTCATCTCTGCCACCCCTCAAG    60
Db      3451 GGCCAGTCAGTGACACTGGCCCTGCCAGGTGTCAGCCAGCAGCGTGCTCATCTCTGCCACCCCTCAAG    3510

Qy      61   AGCAAAGACGGAGCCCCCTGGAGAGCAGCAGCCGTGTCTCATCTCTGCCACCCCTCAAG    120
Db     3511 AGCAAAGACGGAGCCCCCTGGAGAGCAGCAGCGTGCTCATCTCTGCCACCCCTCAAG    3570

Qy     121   AACTTCCAGCTTTCTGACCATCTCTGGTGTGTGGTGTGAGGACCTTGGGTGTGTACACCTGC    180
Db     3571 AACTTCCAGCTTTCTGACCATCTCTGGTGTGTGGTGTGAGGACCTTGGGTGTGTACACCTGC    3630

Qy     181   AGCGTGAGCAATCGCGTGGGAGCAGTGACACCAACGGGCGTCTCTCCGAAGGCAGAGCGC    240
Db     3631 AGCGTGAGCAATCGCGTGGGAGCAGTGACACCAACGGGCGTCTCTCCGAAGGCAGAGCGC    3690

Qy     241   CCTCATCTTTCGCCATGCCCAGGATATCGGGGAGGTGTACGGCGGATGGGGTGTCTGCTGTCTC    300

```

Db 3691 CCCTCATCTTCGCGCATGCCCGGATATCGGGAGGTGTACGGGATGGGTGCTGCTGTC 3750
Qy 301 TGGAAAGCCGCTGGAAATCCTACGGCCCTGTGTACCTACATTTGTGAGTGCAGCTTGAAGGC 360
Db 3751 TGGAAAGCCGCTGGAAATCCTACGGCCCTGTGTACCTACATTTGTGAGTGCAGCTTGAAGGC 3810
Qy 361 GGCAGCTGGACCAACACTGGCCCTCGACATCTTTGACCTGTGCTACCTGACCAAGCTC 420
Db 3811 GGCAGCTGGACCAACACTGGCCCTCGACATCTTTGACCTGTGCTACCTGACCAAGCTC 3870
Qy 421 TCCCGGGGTGCACCTACACCTTCGCGACGCGATGTGTACGAAGGAGGAGGATGGGTCCC 480
Db 3871 TCCCGGGGTGCACCTACACCTTCGCGACGCGATGTGTACGAAGGAGGAGGATGGGTCCC 3930
Qy 481 TACAGCAGCCCTCGGAGCAAGTCTCTCGGAGGGGCCAGCCACCTGCGCTCTGAGGAG 540
Db 3931 TACAGCAGCCCTCGGAGCAAGTCTCTCGGAGGGGCCAGCCACCTGCGCTCTGAGGAG 3990
Qy 541 GAGAGCCAGGGGCGGTGAGCCCAACCCCTGCGCAGCAAAAGACCTTCGCAATCCAGACA 600
Db 3991 GAGAGCCAGGGGCGGTGAGCCCAACCCCTGCGCAGCAAAAGACCTTCGCAATCCAGACA 4050
Qy 601 CAGATCCAGAGGGGCGGCTTCAGGGTGTGTGGGCAATGCTGGGAGAGGCCAGCGGGCGG 660
Db 4051 CAGATCCAGAGGGGCGGCTTCAGGGTGTGTGGGCAATGCTGGGAGAGGCCAGCGGGCGG 4110
Qy 661 GCGTGGCGGCGCAAGATCATCCCTACACCCCAAGGACCAAGACAGCAGCAGTGTGCGGAA 720
Db 4111 GCGTGGCGGCGCAAGATCATCCCTACACCCCAAGGACCAAGACAGCAGCAGTGTGCGGAA 4170
Qy 721 TACGAGGCCCTCAGGGGCTGCGCCACCGCACCTTGGCCAGCTGCGCAGCAGCTACCTC 780
Db 4171 TACGAGGCCCTCAGGGGCTGCGCCACCGCACCTTGGCCAGCTGCGCAGCAGCTACCTC 4230
Qy 781 AGCCCCGGCACCTGGTCTCATCTTGGAGCTGTCTTGGGCCGAGCTGCTCCCTGTC 840
Db 4231 AGCCCCGGCACCTGGTCTCATCTTGGAGCTGTCTTGGGCCGAGCTGCTCCCTGTC 4290
Qy 841 CTGGCCGAGAGGGCCCTCTACTCAGATCCGAGGTGAAGACTACCTGTGCGAGATGTTG 900
Db 4291 CTGGCCGAGAGGGCCCTCTACTCAGATCCGAGGTGAAGACTACCTGTGCGAGATGTTG 4350
Qy 901 AGTGCCACCCAGTACTCTGACAAACAGCACATCTCTGACCTGGACCTGAGGTCCGAGAAC 960
Db 4351 AGTGCCACCCAGTACTCTGACAAACAGCACATCTCTGACCTGGACCTGAGGTCCGAGAAC 4410
Qy 961 ATGATCATCCGGAATACAACTGCTCAAGGTGCTGAGCCT 1001
Db 4411 ATGATCATCCGGAATACAACTGCTCAAGGTGCTGAGCCT 4451

RESULT 10
US-09-858-664A-1
; Sequence 1, Application US/09858664A
; Patent No. US2002072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-858-664A-1

Query Match 99.8%; Score 999.4; DB 9; Length 5207;
Best Local Similarity 99.9%; Pred. No. 6.3e-261;
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GGCAGTCACTGACATGGCTCCAGGTGTACGCCAGCAGCTGCCCCAGGCCACCTGG 60
Db 3628 GGCAGTCACTGACATGGCTCCAGGTGTACGCCAGCAGCTGCCCCAGGCCACCTGG 3687
Qy 61 AGCAAGACGGAGCCCCCTGGAGAGCAGCAGCCGTGTCTCATCTCTGCGACCCCTCAAG 120
Db 3688 AGCAAGACGGAGCCCCCTGGAGAGCAGCAGCCGTGTCTCATCTCTGCGACCCCTCAAG 3747
Qy 121 AACTTCCAGCTTCTGACCAATCCTTGGTGTGTGGTGTGAGCACTTGGTGTGTACCTGC 180
Db 3748 AACTTCCAGCTTCTGACCAATCCTTGGTGTGTGGTGTGAGCACTTGGTGTGTACCTGC 3807
Qy 181 AGGTGAGCAATGGCTGGGGACAGTGACCAACCGGCGTCTCCGAGAGGAGAGAGCGC 240
Db 3808 AGGTGAGCAATGGCTGGGGACAGTGACCAACCGGCGTCTCCGAGAGGAGAGAGCGC 3867
Qy 241 CCCTCATCTTCGCGCATGCCCGGATATCGGGAGGTGTACCGGATGGGGTGTCTGCTGTC 300
Db 3868 CCCTCATCTTCGCGCATGCCCGGATATCGGGAGGTGTACCGGATGGGGTGTCTGCTGTC 3927
Qy 301 TGGAAAGCCCGTGGAAATCCTACGGCCCTGTGACCTTACATTTGTGAGTGCAGCTTGAAGGC 360
Db 3928 TGGAAAGCCCGTGGAAATCCTACGGCCCTGTGACCTTACATTTGTGAGTGCAGCTTGAAGGC 3987
Qy 361 GGCAGTGGACCACTGCGCTCCGACATCTTTGACCTGCTGCTACCTGACCGCAAGCTC 420
Db 3988 GGCAGTGGACCACTGCGCTCCGACATCTTTGACCTGCTGCTACCTGACCGCAAGCTC 4047
Qy 421 TCCCGGGTGGCACCTACACCTTCGCGACGCGATGTGTACGAAGAGGAGGATGGGTCCC 480
Db 4048 TCCCGGGTGGCACCTACACCTTCGCGACGCGATGTGTACGAAGAGGAGGATGGGTCCC 4107
Qy 481 TACAGCAGCCCTCGGAGCAAGTCTCTTGGAGGGGCCAGCCACCTGGCCCTTGAGGAG 540
Db 4108 TACAGCAGCCCTCGGAGCAAGTCTCTTGGAGGGGCCAGCCACCTGGCCCTTGAGGAG 4167
Qy 541 GAGAGCCAGGGGCGGTGAGCCCAACCCCTGCGCAGCAAAAGACCTTCGCAATCCAGACA 600
Db 4168 GAGAGCCAGGGGCGGTGAGCCCAACCCCTGCGCAGCAAAAGACCTTCGCAATCCAGACA 4227
Qy 601 CAGATCCAGAGGGGCGGCTTCTGAGGTGTGGGCAATGCTGGGAGAGGCCAGCGGGCGG 660
Db 4228 CAGATCCAGAGGGGCGGCTTCTGAGGTGTGGGCAATGCTGGGAGAGGCCAGCGGGCGG 4287
Qy 661 GCGTGGCCGCAAGATCATCCCTTACCAACCCCAAGGACAGCAGCTGCTGGCGAA 720
Db 4288 GCGTGGCCGCAAGATCATCCCTTACCAACCCCAAGGACAGCAGCTGCTGGCGAA 4347
Qy 721 TACGAGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCAGCTGCGCAGCAGCTTACCTC 780
Db 4348 TACGAGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCAGCTGCGCAGCAGCTTACCTC 4407
Qy 781 AGCCCCGGCACCTGGTCTCATCTTGGAGCTGTCTTGGGCCGAGCTGCTCCCTGTC 840
Db 4408 AGCCCCGGCACCTGGTCTCATCTTGGAGCTGTCTTGGGCCGAGCTGCTCCCTGTC 4467
Qy 841 CTGGCCGAGAGGGCCCTCTACTCAGATCCGAGGTGAAGACTACCTGTGCGAGATGTTG 900
Db 4468 CTGGCCGAGAGGGCCCTCTACTCAGATCCGAGGTGAAGACTACCTGTGCGAGATGTTG 4527
Qy 901 AGTGCCACCCAGTACTGCAACCAAGCACATCTGCACTGGACCTGAGGTCCGAGAAC 960
Db 4528 AGTGCCACCCAGTACTGCAACCAAGCACATCTGCACTGGACCTGAGGTCCGAGAAC 4587
Qy 961 ATGATCATCCGGAATACAACTGCTCAAGGTGCTGAGCCT 1001
Db 4588 ATGATCATCCGGAATACAACTGCTCAAGGTGCTGAGCCT 4628

Qy 301 TGGAGGCCGCTGGGAATCCTACGGCCCTGTGACCTACATTTGTGCAGTGCAGCTAGAAGGC 360
Db |||||
Qy 3928 TGGAGGCCGCTGGGAATCCTACGGCCCTGTGACCTACATTTGTGCAGTGCAGCTAGAAGGC 3987
Db |||||
Qy 361 GGCAGCTGGACACACACTGGCCCTCCGACATCTTTGACTGTGCTCTACCTGACCCAGCAAGCTC 420
Db |||||
Qy 3988 GGCAGCTGGACACACACTGGCCCTCCGACATCTTTGACTGTGCTCTACCTGACCCAGCAAGCTC 4047
Db |||||
Qy 421 TCCCGGGGTGGACCTACACCTTCGGACGCGATGTGTGACAAAGGAGGAATGGTCCC 480
Db |||||
Qy 4048 TCCCGGGGTGGACCTACACCTTCGGACGCGATGTGTGACAAAGGAGGAATGGTCCC 4107
Db |||||
Qy 481 TACAGACGCCCTCGGAGCAAGTCTCTCGGAGGGGCCAGCCACTTGGAGAG 540
Db |||||
Qy 4108 TACAGACGCCCTCGGAGCAAGTCTCTCGGAGGGGCCAGCCACTTGGAGAG 4167
Db |||||
Qy 541 GAGAGCCAGGGGGGTGAGCCAAACCCCTGCCAGCACAAGAGCTTGGCATTCAGACACA 600
Db |||||
Qy 4168 GAGAGCCAGGGGGGTGAGCCAAACCCCTGCCAGCACAAGAGCTTGGCATTCAGACACA 4227
Db |||||
Qy 601 CAGATCCAGAGGGGCCGCTTCAGCGTGTGTGGGCAATGTCTGGGAGAGGCCAGCGGGCGG 660
Db |||||
Qy 4228 CAGATCCAGAGGGGCCGCTTCAGCGTGTGTGGGCAATGTCTGGGAGAGGCCAGCGGGCGG 4287
Db |||||
Qy 661 GCGCTGGCCGCAAGATCATCCCTTACCACCCCAAGAGCAAGAGCAGCAGTGTCTGCGGAA 720
Db |||||
Qy 4288 GCGCTGGCCGCAAGATCATCCCTTACCACCCCAAGAGCAAGAGCAGCAGTGTCTGCGGAA 4347
Db |||||
Qy 721 TACGAGGCCCTCAAGGGCTCGGGCACCCGACCTGGGCCAGCTGGCCAGCTACCTC 780
Db |||||
Qy 4348 TACGAGGCCCTCAAGGGCTCGGGCACCCGACCTGGGCCAGCTGGCCAGCTACCTC 4407
Db |||||
Qy 781 AGCCCCGGCACCTTGTGCTCATCTTGAGGTGTGTCTCTGGGCCGAGCTGCTCCCTGCG 840
Db |||||
Qy 4408 AGCCCCGGCACCTTGTGCTCATCTTGAGGTGTGTCTCTGGGCCGAGCTGCTCCCTGCG 4467
Db |||||
Qy 841 CTGGCCGAGAGGGCCCTCTACTCAGAATCCGAGGTGAAGACTACCTTGTGCGAGATGTTG 900
Db |||||
Qy 4468 CTGGCCGAGAGGGCCCTCTACTCAGAATCCGAGGTGAAGACTACCTTGTGCGAGATGTTG 4527
Db |||||
Qy 901 AGTGCCACCCAGTACCTGCACAAACAGCACATCTGACCTGGAGCTGAGTCCGAGNAC 960
Db |||||
Qy 4528 AGTGCCACCCAGTACCTGCACAAACAGCACATCTGCACTGGACCTGAGTCCGAGAAC 4587
Db |||||
Qy 961 ATGATCATCCCGAATACAACTGCTCAAGGTCTGGGACCT 1001
Db |||||
Qy 4588 ATGATCATCCCGAATACAACTGCTCAAGGTCTGGGACCT 4628
Db |||||

RESULT 13

US-10-697-263-1
; Sequence 1, Application US/10697263
; Publication No. US20040063142A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV2
; CURRENT APPLICATION NUMBER: US/10/697,263
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 10/274, 978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-697-263-1
Query Match 99.8%; Score 999.4; DB 17; Length 5207;
Best Local Similarity 99.9%; Pred. No. 6.3e-261;
Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GGCCAGTCAGTGACACTGGGCTGCCAGGTGTCCAGCCAGCAGCTGCCCCAGGCCACCTGG 60
Db |||||
Qy 3628 GGCCAGTCAGTGACACTGGGCTGCCAGGTGTCCAGCCAGCAGCTGCCCCAGGCCACCTGG 3687
Db |||||
Qy 61 AGCAAAACAGGAGCCCTCCCTGGAGAGCAGCAGCGTGTCTCTATCTCTGCCACACCTCAAG 120
Db |||||
Qy 3688 AGCAAAACAGGAGCCCTCCCTGGAGAGCAGCAGCGTGTCTCTATCTCTGCCACACCTCAAG 3747
Db |||||
Qy 121 AACTTCCAGCTTCTGACATCTCCTGGTGTGTGCTGAGGACCTGGGTGTGTACACCTGC 180
Db |||||
Qy 3748 AACTTCCAGCTTCTGACATCTCCTGGTGTGTGCTGAGGACCTGGGTGTGTACACCTGC 3807
Db |||||
Qy 181 AGCGTGAAGCAATGCGCTGGGGACAGTGACACACCGGGCGTCTCCGGAAGGACAGAGCGC 240
Db |||||
Qy 3808 AGCGTGAAGCAATGCGCTGGGGACAGTGACACACCGGGCGTCTCCGGAAGGACAGAGCGC 3867
Db |||||
Qy 241 CCCTCATCTTTCGCCATGCCCGGATATCGGGAGGTGTACCGGATGGGTCTCTGCTGTC 300
Db |||||
Qy 3868 CCCTCATCTTTCGCCATGCCCGGATATCGGGAGGTGTACCGGATGGGTCTCTGCTGTC 3927
Db |||||
Qy 301 TGGAGGCCGCTGGGAATCCTACGGCCCTGTGACCTACATTTGTGCAGTGCAGCTAGAAGGC 360
Db |||||
Qy 3928 TGGAGGCCGCTGGGAATCCTACGGCCCTGTGACCTACATTTGTGCAGTGCAGCTAGAAGGC 3987
Db |||||
Qy 361 GGCAGCTGGACACACACTGGGCTCCGACATCTTTGACTGTGCTCTACCTGACCCAGCAAGCTC 420
Db |||||
Qy 3988 GGCAGCTGGACACACACTGGGCTCCGACATCTTTGACTGTGCTCTACCTGACCCAGCAAGCTC 4047
Db |||||
Qy 421 TCCCGGGGTGGACCTACACCTTCGGACGCGATGTGTGACAAAGGAGGAATGGTCCC 480
Db |||||
Qy 4048 TCCCGGGGTGGACCTACACCTTCGGACGCGATGTGTGACAAAGGAGGAATGGTCCC 4107
Db |||||
Qy 481 TACAGACGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACTTGGAGAG 540
Db |||||
Qy 4108 TACAGACGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACTTGGAGAG 4167
Db |||||
Qy 541 GAGAGCCAGGGGGGTGAGCCAAACCCCTGCCAGCACAAGAGCTTGGCATTCAGACACA 600
Db |||||
Qy 4168 GAGAGCCAGGGGGGTGAGCCAAACCCCTGCCAGCACAAGAGCTTGGCATTCAGACACA 4227
Db |||||
Qy 601 CAGATCCAGAGGGGCCGCTTCAGCGTGTGTGGGCAATGTCTGGGAGAGGCCAGCGGGCGG 660
Db |||||
Qy 4228 CAGATCCAGAGGGGGCCGCTTCAGCGTGTGTGGGCAATGTCTGGGAGAGGCCAGCGGGCGG 4287
Db |||||
Qy 661 GCGCTGGCCGCAAGATCATCCCTTACCACCCCAAGAGCAAGAGCAGCAGTGTCTGCGGAA 720
Db |||||
Qy 4288 GCGCTGGCCGCAAGATCATCCCTTACCACCCCAAGAGCAAGAGCAGCAGTGTCTGCGGAA 4347
Db |||||
Qy 721 TACGAGGCCCTCAAGGGCTCGGGCACCCGACCTGGGCCAGCTGGCCAGCTACCTC 780
Db |||||
Qy 4348 TACGAGGCCCTCAAGGGCTCGGGCACCCGACCTGGGCCAGCTGGCCAGCTACCTC 4407
Db |||||
Qy 781 AGCCCCGGCACCTTGTGCTCATCTTGAGGTGTGTCTCTGGGCCGAGCTGCTCCCTGCG 840
Db |||||
Qy 4408 AGCCCCGGCACCTTGTGCTCATCTTGAGGTGTGTCTCTGGGCCGAGCTGCTCCCTGCG 4467
Db |||||
Qy 841 CTGGCCGAGAGGGCCCTCTACTCAGAATCCGAGGTGAAGACTACCTTGTGCGAGATGTTG 900
Db |||||
Qy 4468 CTGGCCGAGAGGGCCCTCTACTCAGAATCCGAGGTGAAGACTACCTTGTGCGAGATGTTG 4527
Db |||||
Qy 901 AGTGCCACCCAGTACCTGCACAAACAGCACATCTGACCTGGAGCTGAGTCCGAGNAC 960
Db |||||
Qy 4528 AGTGCCACCCAGTACCTGCACAAACAGCACATCTGCACTGGACCTGAGTCCGAGAAC 4587
Db |||||
Qy 961 ATGATCATCCCGAATACAACTGCTCAAGGTCTGGGACCT 1001
Db |||||
Qy 4588 ATGATCATCCCGAATACAACTGCTCAAGGTCTGGGACCT 4628
Db |||||

RESULT 14

US-10-697-263-3

; Sequence 3, Application US/10697263

; Publication No. US20040063142A1

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui, et al

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL000927-CIP-DIV2

; CURRENT APPLICATION NUMBER: US/10/697,263

; PRIOR FILING DATE: 2003-10-31

; PRIOR APPLICATION NUMBER: 10/274,978

; PRIOR FILING DATE: 2002-10-22

; PRIOR APPLICATION NUMBER: 09/858,664

; PRIOR FILING DATE: 2001-05-17

; PRIOR APPLICATION NUMBER: 09/711,134

; PRIOR FILING DATE: 2000-11-14

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 5207

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-697-263-3

Query Match 99.8%; Score 999.4; DB 17; Length 5207;

Best Local Similarity 99.9%; Pred. No. 6.3e-261;

Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCCAGTCAGTGACACTGGCTGCAGGTGTACGCCAGCCAGCTGCCAGGCCACCTGG 60

DB 3628 GGCCAGTCAGTGACACTGGCTGCAGGTGTACGCCAGCCAGCTGCCAGGCCACCTGG 3687

QY 61 AGCAAGACGGAGCCCTCGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 120

DB 3688 AGCAAGACGGAGCCCTCGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 3747

QY 121 AACTTCCAGCTTCTGACCATCTGTGGTGTGGTGTGAGGACCTGGGTGTGTACACCTGC 180

DB 3748 AACTTCCAGCTTCTGACCATCTGTGGTGTGGTGTGAGGACCTGGGTGTGTACACCTGC 3807

QY 181 AGCTGAGCAATCGCTGGGACAGTGACACACCGGGGTCTCTCCGAAAGCAGAGCGC 240

DB 3808 AGCTGAGCAATCGCTGGGACAGTGACACACCGGGGTCTCTCCGAAAGCAGAGCGC 3867

QY 241 CCCTCATCTTCGCCATGCCGATATCGGGAGGTGTACGGGATGGGTGTCTGGTTC 300

DB 3868 CCCTCATCTTCGCCATGCCGATATCGGGAGGTGTACGGGATGGGTGTCTGGTTC 3927

QY 301 TGGAGCCCGTGGAACTCTAGCCCTGTGACCTATGTGCACTGTCAGCTCAGAGGC 360

DB 3928 TGGAGCCCGTGGAACTCTAGCCCTGTGACCTATGTGCACTGTCAGCTCAGAGGC 3987

QY 361 GGCAGCTGGACACACTGGCTTCGACATCTTTGACTGTCTACCTGACCAAGCTC 420

DB 3988 GGCAGCTGGACACACTGGCTTCGACATCTTTGACTGTCTACCTGACCAAGCTC 4047

QY 421 TCCCGGGTGGACCTTACACCTTCGACAGGATGTGTACCAAGCAGGAAATGGGTCC 480

DB 4048 TCCCGGGTGGACCTTACACCTTCGACAGGATGTGTACCAAGCAGGAAATGGGTCC 4107

QY 481 TACAGCAGCCCTCGAGCAAGTCTCTCGGAGGGCCAGCAGCTCGGCTCTCAGGAG 540

DB 4108 TACAGCAGCCCTCGAGCAAGTCTCTCGGAGGGCCAGCAGCTCGGCTCTCAGGAG 4167

QY 541 GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCCCAGCACAAAGACCTTCGCAATTCAGACA 600

DB 4168 GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCCCAGCACAAAGACCTTCGCAATTCAGACA 4227

QY 601 CAGATCCAGAGGGGCGCTTCAGCGTGTGTGGGCAATCTGTGGAGAGGGCCAGCGGGCGG 660

Db 4228 CAGATCCAGAGGGGCGCTTCAGCGTGTGTGGGCAATCTGTGGAGAGGCCAGCGGGCGG 4287

QY 661 GCGTGGCGGCAAGATCATCCCTTACCAACCCCAAGAGCAAGAGCAGTGTCTGGCGGAA 720

Db 4288 GCGTGGCGGCAAGATCATCCCTTACCAACCCCAAGAGCAAGAGCAGTGTCTGGCGGAA 4347

QY 721 TACGAGGCGCTCAAGGGCTGGCCACCGCACCTGGGCCAGCTGCAGCAGCCTACCTC 780

Db 4348 TACGAGGCGCTCAAGGGCTGGCCACCGCACCTGGGCCAGCTGCAGCAGCCTACCTC 4407

QY 781 AGCCCCCGGCACTGGTGTCTCATCTTGGAGCTGTCTCTGGGCCGAGCTGTCTCCCTGC 840

Db 4408 AGCCCCCGGCACTGGTGTCTCATCTTGGAGCTGTCTCTGGGCCGAGCTGTCTCCCTGC 4467

QY 841 CTGGCCGAGAGGGCTCTCTACTCAGAAATCCAGAGTGAAAGCACTACCTGTGCGAGATGTG 900

Db 4468 CTGGCCGAGAGGGCTCTCTACTCAGAAATCCAGAGTGAAAGCACTACCTGTGCGAGATGTG 4527

QY 901 AGTGCCACCCAGTACCTGCACAAACAGCAGCATCTGCACCTGGACCTGAGGTCGAGAAC 960

Db 4528 AGTGCCACCCAGTACCTGCACAAACAGCAGCATCTGCACCTGGACCTGAGGTCGAGAAC 4587

QY 961 ATGATCATCACCGAATACAACTGTCTCAAGGTCTGTGGACCT 1001

Db 4588 ATGATCATCACCGAATACAACTGTCTCAAGGTCTGTGGACCT 4628

RESULT 15

US-10-921-168-1

; Sequence 1, Application US/10921168

; Publication No. US20050003446A1

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui, et al

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL000927-CIP-DIV3

; CURRENT APPLICATION NUMBER: US/10/921,168

; PRIOR FILING DATE: 2004-08-19

; PRIOR APPLICATION NUMBER: 10/274,978

; PRIOR FILING DATE: 2002-10-22

; PRIOR APPLICATION NUMBER: 09/858,664

; PRIOR FILING DATE: 2001-05-17

; PRIOR APPLICATION NUMBER: 09/711,134

; PRIOR FILING DATE: 2000-11-14

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 5207

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-921-168-1

Query Match 99.8%; Score 999.4; DB 18; Length 5207;

Best Local Similarity 99.9%; Pred. No. 6.3e-261;

Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGCCAGTCAGTGACACTGGCTGCAGGTGTGAGGACCTGGGTGTGTACACCTGG 60

DB 3628 GGCCAGTCAGTGACACTGGCTGCAGGTGTGAGGACCTGGGTGTGTACACCTGG 3687

QY 61 AGCAAGACGGAGCCCTCGAGAGCAGCAGCGTGTCTCATCTTCGCCACCTCAAG 120

DB 3688 AGCAAGACGGAGCCCTCGAGAGCAGCAGCGTGTCTCATCTTCGCCACCTCAAG 3747

QY 121 AACTTCCAGCTTCTGACCATCTGTGGTGTGGTGTGAGGACCTGGGTGTGTACACCTGC 180

DB 3748 AACTTCCAGCTTCTGACCATCTGTGGTGTGGTGTGAGGACCTGGGTGTGTACACCTGC 3807

QY 181 AGCTGAGCAATCGCTGGGACAGTGACACCAAGCGGCTCTCCGAAAGCAGAGCGC 240

DB 3808 AGCTGAGCAATCGCTGGGACAGTGACACCAAGCGGCTCTCCGAAAGCAGAGCGC 3867

```
Qy 241 CCCTCATCTTGGCCATGCCGGATATCGGGAGGTGTACGGGATGGGTGCTGCTGTC 300
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
3868 CCCTCATCTTGGCCATGCCGGATATCGGGAGGTGTACGGGATGGGTGCTGCTGTC 3927
Qy 301 TGAAGCCCGTGGAAATCCTACGGCCCTGTGACCTTACATTGTGCAGTGCAGCCTAGAAGGC 360
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
3928 TGAAGCCCGTGGAAATCCTACGGCCCTGTGACCTTACATTGTGCAGTGCAGCCTAGAAGGC 3987
Qy 361 GGCAGCTGGACCACTGGCCCTCGACATCTTTGACTGCTGTACTCTGACCAAGCTC 420
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
3988 GGCAGCTGGACCACTGGCCCTCGACATCTTTGACTGCTGTACTCTGACCAAGCTC 4047
Qy 421 TCCCGGGGTGGACCTTACACCTTCCGCAAGCATGTGTACGAAGGAGGAATGGGTCCC 480
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4048 TCCCGGGGTGGACCTTACACCTTCCGCAAGCATGTGTACGAAGGAGGAATGGGTCCC 4107
Qy 481 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACTGGGCCCTCTGAGGAG 540
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4108 TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACTGGGCCCTCTGAGGAG 4167
Qy 541 GAGAGCCAGGGGGGTGAGCCCAACCCCTGCCAGCACAAGACCTTGGCATTCAGACA 600
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4168 GAGAGCCAGGGGGGTGAGCCCAACCCCTGCCAGCACAAGACCTTGGCATTCAGACA 4227
Qy 601 CAGATCCAGAGGGGCGCTTCCAGCGTGGTGGGCAATGTCTGGGAGAGGCCAGCGGGCGG 660
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4228 CAGATCCAGAGGGGCGCTTCCAGCGTGGTGGGCAATGTCTGGGAGAGGCCAGCGGGCGG 4287
Qy 661 GCGTGGCCGCAAGATCATCCCTTACCAACCCCAAGGACAAGACAGTGTCTGGCGAA 720
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4288 GCGTGGCCGCAAGATCATCCCTTACCAACCCCAAGGACAAGACAGTGTCTGGCGAA 4347
Qy 721 TACGAGGCCCTCAAGGGCTGGCCACCCGACCTGGCCAGCTGCACGACGCTACCTC 780
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4348 TACGAGGCCCTCAAGGGCTGGCCACCCGACCTGGCCAGCTGCACGACGCTACCTC 4407
Qy 781 AGCCCCGGCACCTGGTGTCTCATCTTGAGGTGTGTCTCTGGGCCCGAGCTGCTCCCTGC 840
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4408 AGCCCCGGCACCTGGTGTCTCATCTTGAGGTGTGTCTCTGGGCCCGAGCTGCTCCCTGC 4467
Qy 841 CTGGCCGAGAGGGCCCTCCTACTCAGAAATCCAGAGTGAAGGACTACCTGTGGCAGATGTTG 900
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4468 CTGGCCGAGAGGGCCCTCCTACTCAGAAATCCAGAGTGAAGGACTACCTGTGGCAGATGTTG 4527
Qy 901 AGTGCCACCCAGTACTCGCAACCCAGCACATCTTGACCTGGACCTGAGGTCCGAGAAC 960
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4528 AGTGCCACCCAGTACTCGCAACCCAGCACATCTTGACCTGGACCTGAGGTCCGAGAAC 4587
Qy 961 ATGATCATACCGAATACAACCTGCTCAAGTCTGAGGACCT 1001
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
4588 ATGATCATACCGAATACAACCTGCTCAAGTCTGAGGACCT 4628
```

Search completed: March 21, 2005, 16:07:45
Job time : 633.052 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 16:50:48 ; Search time 3663.47 Seconds
(without alignment)
10400.615 Million cell updates/sec

Title: US-10-077-130-4_COPY_3120_4120
Perfect score: 1001
Sequence: 1 aagatgggaagaagctgag.....tcagagcccaaggcgtgtt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_hic:*
- 4: gb_est3:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_gse1:*
- 9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	428.2	42.8	553	7	CN302349 170005326
2	424.2	42.4	621	7	CF179218 813316 MA
3	422.6	42.2	632	7	CF180572 815696 MA
4	377.6	37.7	538	7	CN302346 170005326
5	254.6	25.4	591	5	BU439885 604144795
6	215.4	21.5	1142	4	BG680164 602626828
7	212.2	21.2	401	2	AW898791 CMO-NN007
8	208	20.8	786	6	CD095959 AGENCOURT
9	207.6	20.7	522	8	AQ595824 HS_2132 B
10	194.8	19.5	658	8	AZ383233 1M0140L17
11	193.8	19.4	236	7	CV426951 RC6-EN008
12	193.8	19.4	574	1	AI595154 ml39h10.y
13	161.2	16.1	588	5	BU436399 604145537
14	148.8	14.9	343	2	BE159678 MR0-HT040
15	140.8	14.1	4130	3	AK048144 Mus muscu
16	140.8	14.1	5558	3	AK029863 Mus muscu
17	140.6	14.0	373	1	AA061264 ml39h10.y
18	139.8	14.0	631	4	EG793386 UTSW_SMI1
19	132.2	13.2	397	8	BZ837763 CH240_248
20	127.8	12.8	1192	8	CC247513 CH261-127
21	123.2	12.3	584	7	CK830169 Fr Rev_03
22	123.2	12.3	2066	9	AV399870 Homo sapi
23	123.2	12.3	2274	3	CR590353 full-leng
24	120.4	12.0	472	9	CE495069 tigr-gss-

25	118.2	11.8	179	4	BG989628
26	113.2	11.3	980	5	BX343487
27	108.6	10.8	770	4	BG819323
28	107.8	10.8	2066	9	AY399872 Mus muscu
29	106.2	10.6	1085	4	BM546177 AGENCOURT
30	104	10.4	724	1	AU141131
31	102.8	10.3	569	4	BM254228
32	102.4	10.2	655	1	AU119815
33	102.2	10.2	949	5	BQ645083
34	101	10.1	612	5	BU947303
35	100.4	10.0	886	3	CR607106
36	100.4	10.0	886	5	BU538739
37	100.4	10.0	886	5	BX377136
38	99	9.9	796	5	BX381923
39	98.8	9.9	205	7	CV419760
40	98.2	9.8	1009	5	BQ653435
41	98	9.8	1098	4	BM543727
42	97.6	9.8	575	5	BU072543
43	97	9.7	630	7	CN344412
44	96.6	9.7	626	2	AW743136
45	96.6	9.7	960	5	BQ645277

ALIGNMENTS

RESULT 1
CN302349 553 bp mRNA linear EST 16-MAY-2004
LOCUS 17000532627462 GRN_ES Homo sapiens CDNA 5', mRNA sequence.
DEFINITION CN302349
ACCESSION CN302349
VERSION CN302349.1 GI:47318763
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 553)
AUTHORS Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Flsek, G. J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R., Lebkowski, J. and Stanton, L. W.
TITLE Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 553 Std Error: 0.00.

FEATURES
source
1..553
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, cell lines H1, H7, and H9"
/clone_lib="GRN ES"
/note="Toligo dt-primed, full-length enriched cDNA library from undifferentiated hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions"

ORIGIN
Query Match 42.8%; Score 428.2; DB 7; Length 553;
Best Local Similarity 85.9%; Pred. No. 2e-89;
Matches 475; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 338 GGAGCTGGTGTGTCACAGCGGGCAAGCAGATGCTGGGAGTACAGCTGTGAGGCTG 397
DB 1 GAAGCTGTAGTGCAGCAGGTGGGCAAGCAGATGCTGGGAGTACAGCTGCGAGGCTG 60

```
Qy 398 GGGGCCAGAGAGTCTCTTCCACTGCATCATCAGAGCCCAAGGGGGTGTTCGGAAGG 457
Db 61 GGGGCCAGAGAGTCTCTTCAACTGCATCATCAGAGCCCAAGGGCAGTGTTCGCAAGG 120
Qy 458 AGCAGTCAGTCATATGAGGTGCGCTCAGGCTGAGGCGGGGACCACTGCATCTGAGCTGTG 517
Db 121 AGCAGTTGGTGCATATATGAGGTGCGGACTGAGGCGAGGGCGCAGTGCACACTGAGCTGTG 180
Qy 518 AGGTGGCCCAAGCCCAAGACCGAGGTGACGTGGTGTACAGGACGGGAAGAGCTGAGCTCCA 577
Db 181 AGGTGGCCCAAGCCCAAGACAGAGGTGACGTGGTGTACAGGATGGGAAGAGCTGAGCTCCA 240
Qy 578 GCTCAAAAGTACCGATGGAGGTCAAGGGCTGCACACGAAGGCTGTGTAGTCAGCAGAGTGG 637
Db 241 GTTCGAAGTGCATATAGAGCGCGCGGTGTGATCGGCGAGCTGTGTGTCAGAGCGGAG 300
Qy 638 GCAAAGCAGATGCTGGGGAGTACAGCTGCGAGGCTGGGGCCGAGAGTCTCTTTCAAC 697
Db 301 GCCAGCAGATGCTGGGGAGTACACCTGTGAGGCGCGGGGCCAGCGGCTCTCTTCCACC 360
Qy 698 TGCATCATCAGAGCCCAAGGCAAGTGTTCGCCAAGAGCAGTGTGTGTCATATGAGGTGC 757
Db 361 TGGATGTTTCAGAGCCCAAGGCGGTGTTCGAAAGGAGCAGCTGCGCACACAGAGAGTGC 420
Qy 758 GGAAGTGGCAGGGCCAGTGCACACTGAGCTGTGAGGTGGCCCGAGGCCCAAGAGAGG 817
Db 421 AGGCCAGGCGGGGGCCATTGGCCACGCTGAGCTGCGAGGTGGCCCGAGGCCCAAGAGAGG 480
Qy 818 TGACGTGGTACAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCGCATAGAGGCTG 877
Db 481 TGACGTGGTACAGGAGCGGAAGAGCTGAGCTCCAGCTCGAAGTTCGAATGGAGGCTG 540
Qy 878 CGGGGTGCATGGC 890
Db 541 TGGGCTGCACAGC 553
```

```
RESULT 2
CF179218
LOCUS 813316 MARC 3PIG Sus scrofa cDNA 5', mRNA linear EST 28-JUL-2003
DEFINITION 1 (bases 1 to 621)
ACCESSION CF179218
VERSION CF179218.1 GI:33290994
KEYWORDS EST.
```

```
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
Wise,T.A., Nonneman,D.J., Wray,J.E. and Keele,J.W.
A second set of porcine ESTs from a pooled-tissue normalized
library
```

```
Unpublished (2003)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329
plate: SRG8009 row: O column: 4
Seq primer: GTAATACCACTCACTATAGG.
Location/Qualifiers
1. .621
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3PIG"
```

```
FEATURES
source
```

/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI; Library made with RNA pooled from multiple tissues including brain, liver, muscle, placenta/endometrium, ovary, testes, and bone marrow."

```
ORIGIN
Query Match 42.4%; Score 424.2; DB 7; Length 621;
Best Local Similarity 80.2%; Pred. No. 1.8e-88;
Matches 498; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 108 CAGCTGCGAGGCCAGGGGCCAGAGGGTCTCCCTCCGCTGCACATCACAGAGCCCAAGAT 167
Db 1 CAGCTGCGAGGCTGGGGGCCAAAAGGTCCTCTTCCACTGGATGTCACATAGCCCTCAGT 60
Qy 168 GATGTTTCCAAAGAGAGCAGTCAGTGCATATATGAGGTGCGAGGCTGAGGCGGGGGCCAGTGC 227
Db 61 GGTGTTTCCAAAGAAACAGCTGGGCTGCGAGTGAAGTGCAGGCCATGGCAGGGGCCATTGC 120
Qy 228 CATGCTGAGCTGTGAGGTGGCCCGAGGCCAGCGAGGTGACGTGGTACAGAGGATGGGA 287
Db 121 CACGCTGAGCTGTGAGGTGGCCCATGCCCATACGAGAGTGACTTGGTACAGAGGATGGAA 180
Qy 288 GAACTGAGCTCCAGCTCAAAAGTGGGCATGGAGGTCAAAAGGGTGCACACGAGAGGCTGGT 347
Db 181 GAAGCTGAGTTCGAGCTCAAAAGTGCCTGTGGAGGCCAAGGGCTGCAACCGAGGGCTGGT 240
Qy 348 GCTCCACAGCGCGCAAGCAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGGCCAGAG 407
Db 241 GGTGCATCAAGTGGCGAAGGCATATGCTGGGGAGTACAGCTGGCATGCTGGGGGTCAATA 300
Qy 408 AGTCTCTTCCACTGCACATCACAGAGCCCAAGGGGGTGTTCGGAAGGAGCAGTGCAGT 467
Db 301 GGTCTCTTCCACTGGATGTCACATATCCCTCAGTGGTGTTCGCAAGGAGCAGGCATC 360
Qy 468 GCATATGAGCTGCGAGCTGAGGGGGGACCATGCCCATGCTGAGCTGTGAGGTGGGCCCA 527
Db 361 ACGCACTGAAGTCAAGCCCATGGCGGAGCAGTGCACACTGAGCTGTGAGGTGGGCCCA 420
Qy 528 GCCCCAGACGAGGTGACGTGGTACAGGACGGAAGAGCTGAGCTCCAGCTCCAAAGT 587
Db 421 TGCCAGACAGAGTGATGTTACAGGATGGGAAGAGCTGACTTCAAGCTCAAAAGT 480
Qy 588 ACGCATGGAGGTCAAGGGTGCACACGAAGGCTGGTAGTGCAGCAGGTGGGCCAAAGCAGA 647
Db 481 CGCTGTGGAGGCCAATGGCTGCAACCGATGCTGGTGGTGCATCATGCGGGCAATGCATA 540
Qy 648 TGCTGGGGAGTACAGCTGCGAGGCTGGGGCCAGAGAGTCTCTTCAACTGCACATCAC 707
Db 541 TGCTGGGGAGTACAGCTGCGATGCTGGGGGTCAATAATGTCTCTTCCACTGGACATCAC 600
Qy 708 AGAGCCCCAAGGCAGTGTTCG 728
Db 601 AGAGCCCTCAGTGTGTTCG 621
```

```
RESULT 3
CF180572/c
LOCUS 815696 MARC 3PIG Sus scrofa cDNA 3', mRNA linear EST 28-JUL-2003
DEFINITION 1 (bases 1 to 632)
ACCESSION CF180572
VERSION CF180572.1 GI:33292348
KEYWORDS EST.
```

```
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
Wise,T.A., Nonneman,D.J., Wray,J.E. and Keele,J.W.
A second set of porcine ESTs from a pooled-tissue normalized
library
Unpublished (2003)
Contact: Smith TPL
```


Db 421 TGACGAGGATGCCAGGCGACACCGGGAGTATAGCTGCGAGGCCGGGGCCAGCGGC 480

QY 962 TCTCTTCCACCTGGATTTTCAG 985

Db 481 TCTCTTCCAGCTGGACGTGGCAG 504

RESULT 5
BU439885
LOCUS 591 bp mRNA linear EST 29-NOV-2002
DEFINITION 604144795F1 CSEQRN11 Gallus gallus cDNA clone CHEST983m23 5', mRNA sequence.
ACCESSION BU439885
VERSION BU439885.1 GI:25929196
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 591)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
Location/Qualifiers
1..591
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST983m23"
/sex="Male and female"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQRN11"
/note="Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 25.4%; Score 254.6; DB 5; Length 591;
Best Local Similarity 65.7%; Pred. No. 7.8e-49;
Matches 371; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 431 CAGAGCCACAGGGGTTTCGAGAGGAGCAGTGCATATGAGTGCTGAGG 490

Db 1 CAGAGCCAGAGTGTGTATTATCAACAGAGGAGAGGTGCAGAGGCTGAC 60

QY 491 CGGGGACCACTGCCATGCTGAGCTGTGAGGTGGCCAGCCAGACCGAGGTGACGTGGT 550

Db 61 CAACAGAAAATGCCACGCTGAGCTGCGAGGTGGCCAGGAGACGCGAGTGAAGTGGT 120

QY 551 ACAAGACCGGAAGAAGCTGAGCTCCAGCTCAAGCTCAAAAGTACGATGAGAGTCAAGGGCTGCA 610

Db 121 ACAAGGAGGGGAAAACCTGATCACTCCGACCAAGAAGTTTCAGGGTGGAGTTCAGAGGGCAAAAC 180

QY 611 CACGAAGGCTGCTAGTGCAGCAGGTGGCAAAAGCAGATGCTGGGGAGTACAGCTGCCAGG 670

Db 181 TGGCTCGGCTGGTGGTGGCCAGGTGGAGAAGAAGATGCGGGGAGTACACCTGCCGAGG 240

QY 671 CTGGGGCCAGAGAGTCTCTTTCAACTGCACATCAGAGAGCCCAAGGCACTGTTGGCA 730

Db 241 CTGCTGCCAGAAACTGACCTTCAAGATTGATGTACAGAGCCCAAAACCTGTGTTTATAA 300

QY 731 AGGAGCAGTGGTGCATATGAGTGGCGGATGAGGAGGGGCCAGTGCACACTGAGCT 790

Db 301 ACCAGGAGAGGTGCAGAGGGAGGTGAGTGTCTTGGCAGAAAAGTGCCTCCCTCAGCT 360

QY 791 GTGAGGTGGCCCGCCAGCCAGACAGAGTGCCTGCTTCAAGATGGGAAGAGTGCAGCT 850

Db 361 GTGAAGTAGCACAGGATACACCGAAGTAAATGTGTACAGGATGGAAGACTGCTTGCCT 420

QY 851 CCAGTTCGAAGTGGCGATAGAGCTCGGGCTGCATGCGGCACTGCTGCTGCAGCAGG 910

Db 421 CCTCTAGAAAAGTTCAAAATGGAACCTGTGGGCAAAACCCCGCGTCTGTTGTGAGCAGC 480

QY 911 CAGGCCAGGAGTGTGGGGAGTACACCTGTGAGGCTGGGGCCAGCGGCTCTCCTTCC 970

Db 481 TGGAGAAGAAGATGCTGGAGAATACGTGTGAGGCTGCGCGCCAGAGGCTGACCTTCA 540

QY 971 ACCTGGATGTTTCAGAGCCCAAGGC 995

Db 541 AGTTGGAAGCACTGAACAGAGGC 565

RESULT 6
BG680164/c 1142 bp mRNA linear EST 01-MAY-2001
LOCUS 602626828F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4751671 5', mRNA sequence.
DEFINITION BG680164
ACCESSION BG680164
VERSION BG680164.1 GI:13911561
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1142)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM10608 row: h column: 08
High quality sequence stop: 708.
Location/Qualifiers
1..1142
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4751671"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skn4"
/note="Organ: skin; Vector: pCMV-SORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

FEATURES
source
Location/Qualifiers
1..1142
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4751671"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skn4"
/note="Organ: skin; Vector: pCMV-SORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI-CGAP Library."

```
ORIGIN
Query Match      21.5%; Score 215.4; DB 4; Length 1142;
Best Local Similarity 84.1%; Pred. No. 1.3e-39;
Matches 243; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 697 CTGCACATCAGAGAGCCCAAGCGAGTGTTCCTCCAGGAGCAGTGTGTCATAATGAGGTG 756
    |||
Db 322 CTGCTCTCCAGAGAGCCCAAGCGAGTGTTCCTCCAGGAGCAGCAGCAGCAGGAGGTG 263
    |||

Qy 757 CGGACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 816
    |||
Db 262 CAGGCTGAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 203
    |||

Qy 817 GTGACGTGTGTTACAAAGGATGGGAAGAGTCCAGTCCAGTTCGAAAGTCCGATAGAGGCT 876
    |||
Db 202 GTGACTTGGTACAAAGGACGGGAAGAGTCCAGTCCAGTTCGAAAGTCCGATAGAGGCT 143
    |||

Qy 877 GCGGGCTGCATGCGGACGCTGGTGTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 936
    |||
Db 142 GCAGGCTGTACAGGAGGCTGGTGTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 83
    |||

Qy 937 ACCTGTGAGGCTGGGGGCGAGCGGCTCTCTCCACCTGGATGTTTCAG 985
    |||
Db 82 AGCTGTGAGGCTGGGGGCGAGCGGCTCTCTCCACCTGGATGTTTCAG 34
    |||

RESULT 7
AW898791
LOCUS CMO-NN0075-130400-332-f06 NN0075 Homo sapiens cdna, mRNA sequence.
DEFINITION CMO-NN0075-130400-332-f06 NN0075 Homo sapiens cdna, mRNA sequence.
ACCESSION AW898791
VERSION AW898791.1 GI:8062996
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 401)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Negai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=CMO-NN0075-130
400-332-f06&t3=2000-04-13&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 401.
Location/Qualifiers
1..401
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN0075"

FEATURES
source
1..786
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6940598"
/dev_stage="metamorphosis stage 53"
/clone_lib="NICHD XGC Tad1"
/note="Organ: Developing Tadpole; Vector: pDNR-LIB;
Site_1: Sfi; Site_2: Sfi; 5' and 3' adaptors were used in

/note="Organ: nervous normal; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```


84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunngenet@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0140 row: L column: 17
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 658.
Location/Qualifiers
FEATURES
source
1..658
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC1M0140L17"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, P-"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 236)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,P.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. http://www.ludwig.org.br.
FEATURES
source
1..236
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="EN0083"
/note="Organ: lung_normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196.716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
Query Match 19.4%; Score 193.8; DB 7; Length 236;
Best Local Similarity 96.6%; Pred. No. 1,1e-34;
Matches 198; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 664 TCGAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCATCACAGAGCCCAAGCAGTG 723
Db 32 TGCRAAGTTGCAGGCCAGAGAGTCTCTTTCAACTGCATCACAGAGCCCAAGCAGTG 91
QY 724 TTTCGCAAGGAGCAGTTGGTGCATTAATGAGTGGGACTGAGGCGAGGGCCAGTGCCACA 783
Db 92 TTTGCAAGGAGCAGTTGGTGCATTAATGAGTGGGACTGAGGCGAGGGCCAGTGCCACA 151
QY 784 CTGAGCTGTGAGTGGGCCAGCCAGACAGAGGTGAGTGTGTACAGATGGGAAGAG 843
Db 152 CTGAGCTGTGAGTGGGCCAGCCAGACAGAGGTGAGTGTGTACAGATGGGAAGAG 211
QY 844 CTGAGCTCCAGTTCGAAAGTGCACA 868
Db 212 CTGAGCTCCAGTTCGAAAGTGCACA 236
RESULT 12
LOCUS AI595154
DEFINITION ml39h10.v1 Stratagene mouse testis (#937308) Mus musculus cDNA
ACCESSION AI595154
VERSION AI595154.1 GI:4604202
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Query Match 19.5%; Score 194.8; DB 8; Length 658;
Best Local Similarity 81.3%; Pred. No. 7.6e-35;
Matches 226; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 156 AGAGCCCAAGATGATGTTTGCAGGAGCAGTCAGTGCATATAGTGGTGCAGGCTGAGGC 215
Db 182 AGACACCAGGCTGATGTTTGCAGGAGCAGCAGGCGAGGCTGAGGCTGAGGC 241
QY 216 GGGGCCAGTGCATGCTGAGTGTGAGTGGCCAGCCAGCCAGCGAGGTGACGTGTA 275
Db 242 GGGGAACAGTGCACCCCTGAGCTGCGAGTGGCCAGCCAGCCAGACTGAGGTGACATGTT 301
QY 276 CAAGGATGGGAAGAGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGGTCAAGGCGTGAC 335
Db 302 CAAGAGCGGAAGAGCTGAGCTCCAGCTCGAAGTGGCGTGGAGGCTCGGGCTGCTC 361
QY 336 ACGGAGGTGTGTGTCACACAGCGCGGCAAGCAGATGCTGGGGAGTACAGCTGTGAGGC 395
Db 362 CAGGAGGTGTGTGTCAGCAGCGCGGCAAGCGGATGCTGGGGAGTACAGCTGCGAGGC 421
QY 396 TGGGGCCAGAGAGTCTCTTCCACTGCACATACAG 433
Db 422 CGGGGGTCAGAGCTCTCTTCCGCTGACGTGCGAG 459
RESULT 11
LOCUS CV426951
DEFINITION RC6-EN0083-311000-011-H11 EN0083 Homo sapiens cDNA, mRNA sequence.
ACCESSION CV426951
VERSION CV426951.1 GI:52822454
KEYWORDS EST.
SOURCE Homo sapiens (human)

REFERENCE
AUTHORS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 574)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., McCann, R.,
Ritter, B., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., Schurr, R.,
Waterston, R., and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LILNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
MGI:308283
Seq primer: -40RP from Gibco
High quality sequence stop: 418
POLYA=No.

FEATURES

source
1..574
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:514435"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse testis (#937308)"
/note="Organ: testis; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

ORIGIN

Query Match 19.4%; Score 193.8; DB 1; Length 574;
Best Local Similarity 80.9%; Pred. No. 1.3e-34;
Matches 225; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 156 AGAGCCCAAGATGATGTTTGCAGAGCAGTCAGTCATATAGGTGCAGGCTGAGGC 215
Db 465 AGACACCAAGCTGATGTTTGCAGAGGAGCAGCAGCGACCGCAGGTTGAAGGCTGAGGC 406
QY 216 GGGGGCCAGTGCATGCTGAGCTGTGAGGTGGCCCGCCAGCCAGCGAGGTGACGTGGTA 275
Db 405 GGGGAACAGTGCACCTGAGCTGCAGGTGGCAGCCAGCCAGCTGAGGTGACATGTT 346
QY 276 CAAGATCGGGAAGAGCTGAGCTCAGCTCAAAAGTGGGCATGAGGTGCAAGGGTGCAC 335
Db 345 CAAGACGGGGAAGAGCTGAGCTCAGCTCAGAGGTGGCGCTGGAGGCTCGGGCTGCTC 286
QY 336 ACGGAGGCTGTGTGCTGCCACAGGGGGCAAGCAGATGCTGGGAGTACACTCTGAGGC 395
Db 285 CAGGAGGCTGTGTGCTGCAGAGGGGGCAAGCGGATGCTGGGAGTACACTCGAGGC 226
QY 396 TGGGGGGCAGAGAGTCTCTTTCCACCTGCACATCACAG 433
Db 225 CGGGGGGCAGAGAGTCTCTTTCCGCGCTGGAGCTGGCAG 188

RESULT 13
BU436399
LOCUS
DEFINITION
604145537F1 CSEQRBN11 Gallus gallus cdna clone CHEST98519 5', mRNA
sequence.

ACCESSION

BU436399

VERSION

BU436399.1

KEYWORDS

EST

SOURCE

Gallus gallus (chicken)

ORGANISM

Gallus gallus

REFERENCE

1 (bases 1 to 588)

AUTHORS

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A., and Hubbard, S.J.

TITLE

A Comprehensive Collection of Chicken cDNAs

JOURNAL

Curr. Biol. 12 (22), 1965-1969 (2002)

MEDLINE

22335534

PUBMED

12445392

COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

1..588
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST98519"
/sex="Male and female"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQRBN11"
/note="Vector: pBluescript II KS(+); Site 1: EcoRI;
Site 2: NotI; This normalized library was constructed from
1 million independent clones. cDNA synthesis was initiated
using an oligo(dt) primer, using methylated C in the first
strand synthesis reaction. Following this first strand
reaction, double-stranded cDNA was blunted, ligated to
NotI adapters, digested with EcoRI, size-selected, and
cloned into the NotI and EcoRI compatible sites of a
custom modified MCS of the pBluescript (KS+) vector. The
library was normalized in 2 rounds using conditions
adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
Bonaldo et al., Genome Research 6 (1996): 791, except that
a significantly longer reannealing hybridization was
used."

ORIGIN

Query Match 16.1%; Score 161.2; DB 5; Length 588;
Best Local Similarity 67.1%; Pred. No. 5.4e-27;
Matches 243; Conservative 0; Mismatches 118; Indels 1; Gaps 1;
QY 619 CTGCTAGTCAGCAGGTGGGCAAGCAGATGCTGGGAGTACAGCTCGAGGCTGGGGGC 678
Db 1 CTGCTGTGTAGCCAGGTGGGAGAGAAAGATGC-AGGAGTACACCTCGAGGCTGCTGGC 59
QY 679 CAGAGAGTCTCCCTTTCACATGTCACATCAGAGCCCAAGCAGTGTGTCAGAGGACAG 738
Db 60 CAGAAACTGACCTTCAAGATTACTGTCTACAGAGCAGAGGATGCTTTTATCAACAGGAG 119
QY 739 TTGCTGCATAATAGGTGGCGACTGAGGAGGGGCGCAGTCCACACTGACCTGTGAGGTG 798
Db 120 AAGTGCAGAGAGGAGGTGAGGCTGCACCAAGAGAAATGCCAGCTGACCTGCGAGGTG 179
QY 799 GCCCAGGCCCAGACAGAGGTGACGTGGTACAAGGATGGGAAGAGCTGACCTCAGTTCG 858
Db 180 GGGCAGGAGAGAGCCGAGGTGAAGTGGTACAAGAGGGGAGAACTGATCACCTCGAGCAAG 239
QY 859 AAGTGGCATAGAGGTGCGGGCTGCATCGGCAGCTGGTGGTGCAGCAGGAGGCGCAG 918

Db 240 AAGTTGAGGTGAGTACAGAGGCAAACTGCTGGCTGGTGGTGGAGCCAGCGTGGAGAG 299

Qy 919 GCAGATGCTGGGAGTACACCTGTGAGCTGGGGCCAGCGGCTCTCTTCCACCTGGAT 978

Db 300 AGAAATGCGGGGAGTACACCTGCGAGGCTGCTGGCCAGAAAGTGAGCTTCAAGATTACT 359

Qy 979 GT 980

Db 360 GT 361

RESULT 14

LOCUS BE159678 343 bp mRNA linear EST 21-JUN-2000

DEFINITION MRO-HT0407-180400-015-e07 HT0407 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE159678

VERSION BE159678.1 GI:8622399

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Dias Neto, E., Garcia Correa, R., Verlovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-MRO-HT0407-180400-015-e07&t3=2000-04-18&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 343.
Location/Qualifiers
1. .343
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0407"
/note="Organ: head neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORFESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES source

ORIGIN

Query Match 14.9%; Score 148.8; DB 2; Length 343;
Best Local Similarity 88.3%; Pred. No. 3.9e-24;
Matches 173; Conservative 0; Mismatches 22; Indels 1; Gaps 1;

Qy 153 CACAGAGCCCAAGATGATGTTTGCAGAGGAGCAGTGCATATGAGTGGCAGGCTGA 212

Db 228 CCCAGAGCCCAAGAGTGGTGTGTTGCCAAGAGCAGCCACACACAGGGAGTGCAGGCTGA 169

Qy 213 GCGCGGGGCCAGTGCATGCTGAGCTGTGAGGTGGCCCGCCAGCGGAGGTGACGCTG 272

Db 168 GCGCGGGGCCAGTGCACGCTGAGCTGCGAGGTGGCCCGCCAGCAGAGGTGACGCTG 109

Qy 273 GTACAAGGATGGGAAGAGCTGAGCTCCAGCTCAAAAGTGGGCGATGGAGGTCAAAAGGCTG 332

Db 108 GTACAAGGATGGGAAGAGCTGAGTTCAGCTCGAAAGTGGCGCTGGA-GCGCGTGGCTG 50

Qy 333 CACACGGAGGCTGGT 348

Db 49 CACACGGAGGCTGGT 34

RESULT 15

LOCUS AK048144

DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone: C130037H06 product: hypothetical immunoglobulin and major histocompatibility complex domain containing protein, full insert sequence.

ACCESSION AK048144

VERSION AK048144.1 GI:26339169

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 92779253

PUBMED 10349636

REFERENCE

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N.,

Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

FEATURES

Source

1. 4130
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C130037H06"
/db_xref="taxon:10090"
/clone="C130037H06"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
660. 3491

CDS

/note="unlabeled protein product; hypothetical
immunoglobulin and major histocompatibility complex domain
containing protein (InterPro|IPR003006, evidence:
InterPro)
putative"

/codon_start=1

/protein_id="BAC33256.1"

/db_xref="GI:26339170"

/translation="MRGRVTVANVTVKVGPILKRLPKLDVLEGENAVLIVETQAGV
OGWSRDEGLPDTQSCGCHMRALVLPVTRDAGEITFSLGNSRTTLLRVKCVKH
SPGSPVNMVEFGKGNKLVLTWKPPEPPETSYFLRLEROEVGSDWIOCFSEKAG
AVEPVGDTGDFYHFRICTVSEHRSPPHVFNGSAHLVPTARLVSGLEDVQVYDGE
DAVSLDSALIIQGSFWLNGQLQNEPQGVFPCALRYIEQKGLHRLILQAVKHR
DSGALVPCSGVDQSAALITQESSVHILSPQDKVSLTFTTSRVLTLCELSRVDPFA
TWYKDGQVSESSLIIVKTEGRKRLILPEAQVDSGEFECRTGVSAFQVTVQDPP
VHVNPOHVFVHALTSECVRILTCEVDREDTTVHMYKDGQVSESDIIVLENKPHRR
LVLPAAKSDGGEFCQVAGDERAFTVTITDVFNSVIPTPSSEVHVAARLVRLVLTCE
LCRPAEVRVTKDGEVVEVSPALLLEKEDTIRLVLPVQLDESGEYLCEIHDSASF
TITVTEPPVRIIYQDEVTLHVASLECVLTCELSREDAPVRYKDGLEVESEBALVL
QSDGPRRLVLPAAQPEDGFEVCDAGDSDSAFTVTVPPIVHPAARSLDIQFGA
PGHVELRCEVAPAGSVRYKDGLEVEVSDALQIARGPARTLTLPHQPEDAGEVYC
ETRDPAVTVNSLAELPVQFLAPEANPLCVRGEVPLVSCELSRASQAQVFWSHNS
PVQOGEGLERAEGRPRILICIAADLAHTGVITCQSGASPGASISFNQVAELPPVK
LVSELTPLVHEGDDATFQCEVSPDPAEVTWLRNGAVITAGPQLEVMVQNGSSRLTIIR
GCQLKDGTVTVARAGADTSARLHVRGSV"

ORIGIN

Query Match 14.1%; Score 140.8; DB 3; Length 4130;
Best Local Similarity 47.5%; Pred. No. 4.5e-22;
Matches 452; Conservative 0; Mismatches 497; Indels 3; Gaps 1;

Qy 1 AAAGATGGGAAGAAGCTCAGCTCCAGCTTGAAGAGTGATGTAGAGGCCAAAGCGCTGCAGA 60
Db 2217 AAAGATGGGAGGAGGTAGTGAGAGGCCAGCAGCTGCTCCTGGAGAGGAGACCATC 2276
Qy 61 CGGAGGCTGGTGTGCACAGCAGCGAGGCAAGACGATGCCGGGGACTACAGCTCGAGGCC 120
Db 2277 CGCGGCTGGTGTGCTGCTGCTGTCCAGCTTGAGGATTCGGCGAGTACCTGTGTGAAATC 2336
Qy 121 AGGGCCACAGAGGGTCTCCTCTCCGCTGCACATCACAGAGCCCAAGATGATGTTTTCAG 180

Search completed: March 21, 2005, 14:13:33
Job time : 3659.47 secs

Db 2337 CATGATGAGTCGGCTTCTCTTACCATCACCGCTCACAGAGCCCTGTCGCGATCATATAC 2396
Qy 181 GAGCAGTCAGTGCATAATAGGTGTCAGGCTGAGCGGGGGCCAGTGCATCTCAGCTGT 240
Db 2397 CCCAGAGCAGAGGTGACCT---TACGCGCTGAGTTTGAATGTGTGCTCACCTGT 2453
Qy 241 GAGTGGCCAGGCCAGAGCGAGGTGACGTGGTGTACAGGATGGGAAGCAAGCTCAGCTCC 300
Db 2454 GAGCTGTCTAGAGAGGATGCTCCTGTAGCTGGTGTACAGGATGGTTAGAGTGGAGGAG 2513
Qy 301 AGCTCAAAAGTGGCGATGGAGGTCAAGGGTGCACACGAGGCTGGTGTGCCACAGGCG 360
Db 2514 AGTGAAGCCCTGGTGTCTCCAGAGCGATGGGCTCTGTCGCGCTGTGGTGTGGCTGCC 2573
Qy 361 GGCAAAACAGATCTGGGGAGTACAGCTGTGAGGCTGGGGCCAGAGAGTCTCCTCCAC 420
Db 2574 CAGCCAGAGCAGGGGGCGAGTTTGTGTGTGATCTGGGGATGATTGAGCCTTCTTCACT 2633
Qy 421 CTGCACATCACAGAGCCCAAGGGGTGTTCGCAAGGAGCAGTCAATGATGAGGTG 480
Db 2634 GTCACTGTCAAGCTCCACAGAAAGGATTGTGACCCAGCGGCCCGATCCCTGATTG 2693
Qy 481 CAGCTGAGCGGGGGACCACTGCCATCTGAGCTGTGAGTGTGGCCACGCCACAGCGAG 540
Db 2694 CAGTTTGGGGCTCCAGGACACGTGGAGCTACGCTGCGAAGTGGCCCGGCTGGGTCTCAG 2753
Qy 541 GTGACGTGGTACAGGACGAGGAGAGAGCTGAGCTCCAGCTCAAAAGTACGATCGAGGTC 600
Db 2754 GTGGCTGGTACAGGATGGCTTAGAGTAGAGGTGTAGATGGCTGCGAGTGGGTGCT 2813
Qy 601 AAGGGCTGCACAGAAAGGCTGGTAGTGACAGAGTGGGCAAGCAGATGCTGGGGAGTAC 660
Db 2814 GAGGGGCTGCCCCGCACTCTCACCTGCCCCACGCCAGCTGAGGATGCCGGGAGTAT 2873
Qy 661 AGCTGCGAGCTGGGGCCAGAGAGTCTCTTTCACTGCACATCACAGAGCCCAAGGCA 720
Db 2874 GTATGTGAGACCCGAGATGAGGCTGTCACTTCAACGTGAGCTGGGTGAGCTTCCGGTG 2933
Qy 721 GTGTTTCCCAAGGAGCAGTTGGTGCAATAAGTGGCGGACTGAGGCGAGGGGCCAGTGCC 780
Db 2934 CAGTTTCTGGCTCCAGAGGCGAGCCCCAATCCGCTCTGCGTGGTTCTTGGGGAGCCCTG 2993
Qy 781 ACATGAGCTGTGAGTGGGCCAGAGGCCACAGACAGAGTGACGTGGTTACAAGAGTGGAAAG 840
Db 2994 GTGCTGAGCTGTGAGCTGTCCGAGCAAGCGCACAGTGTCTTGGAGCCCAATGGGAGC 3053
Qy 841 AAGCTGAGCTCCAGTTTCAAGTCCGATAGAGGCTGCGGGCTGCATGCGGCGAGCTGGTG 900
Db 3054 CCGGTGAGCAGGGGTGAAGACTAGAGCTCGAGCTGAGGGTCTCTCCAGAAATCTCTGC 3113
Qy 901 GTGAGCAGGCGAGGCCAGGAGATGCTGGGGAGGTACACCTGTGAGGCTGGGG 952
Db 3114 ATCCAGGACAGACCTCGCTCACAAGGTGTCTACACTTGGCCATCTGGGG 3165

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:15:10 ; Search time 581.044 Seconds
(without alignments)
10198.307 Million cell updates/sec

Title: US-10-077-130-4_COPY_22500_23500

Perfect score: 1001

Sequence: 1 ggccagtcagtgacactggc.....ctgtcaagtcgtggacct 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_16Dec04.*

1: geneseqn1980s.*

2: geneseqn1990s.*

3: geneseqn2000s.*

4: geneseqn2001as.*

5: geneseqn2001bs.*

6: geneseqn2002as.*

7: geneseqn2002bs.*

8: geneseqn2003as.*

9: geneseqn2003bs.*

10: geneseqn2003cs.*

11: geneseqn2003ds.*

12: geneseqn2004as.*

13: geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	3225	5	Aad26467 Human kin
2	1001	100.0	4175	6	Aac62285 cDNA enco
3	1001	100.0	4936	4	Aah46304 cDNA enco
4	1001	100.0	5454	6	Aad38865 Human kin
5	1001	100.0	8106	8	Abx11641 Human ser
6	1001	100.0	24120	8	Abx11642 Human ser
7	999.4	99.8	5007	5	Aac62286 cDNA enco
8	999.4	99.8	5207	6	Aal43909 Human kin
9	999.4	99.8	5207	6	Aal43908 Human kin
10	999.4	99.8	7928	5	Aac62287 cDNA enco
11	385	38.5	876	7	Adr41256 Human CD-
12	164.8	16.5	182	3	Aac25654 Human sec
13	151.6	15.1	9930	10	Ade47671 Human NOV
14	151.6	15.1	9930	12	Adj78941 Human NOV
15	150	15.0	7710	4	Aaf44662 Novel pro
16	150	15.0	7710	12	Adi29360 Human MAR
17	148.4	14.8	7789	6	Aad30565 Human kin
18	148.4	14.8	10662	10	Adc99127 Human KPP
19	146.8	14.7	9698	10	Ade47675 Human NOV
20	146.8	14.7	9698	12	Adj78945 Human NOV

21	146.8	14.7	10122	10	AdE47673
22	146.8	14.7	10122	12	ADJ78943
23	140.4	14.0	9807	10	ADB79958
24	105.6	10.5	62805	10	ADB79960
25	97.8	9.8	860	10	ADB79960
26	97.8	9.8	860	12	ADJ78947
27	92	9.2	1365	12	ADJ78947
28	92	9.2	2055	13	ADN60271
29	92	9.2	2055	13	ADR83403
30	92	9.2	2079	4	AAK94258
31	92	9.2	2079	13	ADL30841
32	92	9.2	2079	13	ADQ68012
33	92	9.2	2104	12	ADQ84258
34	92	9.2	2104	13	ADQ85644
35	92	9.2	2104	13	ADQ86740
36	92	9.2	2224	4	AAH16158
37	92	9.2	2224	5	AAH78068
38	92	9.2	2226	10	ADD29770
39	90.4	9.0	757	4	AAK93262
40	90.4	9.0	757	4	AAK91856
41	90.4	9.0	757	12	ADL29689
42	90.4	9.0	757	12	ADL28283
43	90.4	9.0	2105	6	ABV72291
44	90.4	9.0	2105	13	ADN60267
45	90.4	9.0	2105	13	ADN60269

ALIGNMENTS

RESULT 1

AAD26467

ID AAD26467 standard; cDNA; 3225 BP.

AC AAD26467;

DT 26-MAR-2002 (first entry)

DE Human kinase PKIN-20 cDNA.

XX Human; kinase; PKIN-20; cancer; leukaemia; adenocarcinoma; osteoporosis;

XX Immune disorder; atherosclerosis; Crohn's disease; Hodgkin's disease;

XX Acquired Immune Deficiency Syndrome; AIDS; Addison's disease; anaemia;

XX allergy; asthma; adult respiratory distress syndrome; multiple sclerosis;

XX autoimmune thyroiditis; bronchitis; diabetes mellitus; osteoarthritis;

XX Good pasture's syndrome; Graves' disease; pancreatitis; psoriasis;

XX rheumatoid arthritis; ulcerative colitis; cirrhosis; Cushing's syndrome;

XX hepatitis; hypothyroidism; cerebral palsy; cataract; angina pectoris;

XX cardiovascular disease; hypertension; vasculitis; myocarditis; obesity;

XX congestive heart failure; ischaemic heart disease; lung tumour; gout;

XX fatty liver; Niemann-Pick's disease; gene therapy; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

XX CDS 526..3141

XX /*tag= a

XX /product= "Human PKIN-20 protein"

XX WO200196547-A2.

XX 20-DEC-2001.

XX 14-JUN-2001; 2001WO-US019444.

XX 15-JUN-2000; 2000US-0212073P.

XX 23-JUN-2000; 2000US-0213467P.

XX 30-JUN-2000; 2000US-0215651P.

XX 07-JUL-2000; 2000US-0216605P.

XX 13-JUL-2000; 2000US-0218372P.

XX 25-AUG-2000; 2000US-0228056P.

XX (INCY-) INCYTE GENOMICS INC.

PR	16-APR-1999;	99US-0129553P.
XX	(SCIO-) SCIOS INC.	
XX	Zeng W, Stanton L, Kong H;	
XX	WFI; 2001-007013/01.	
DR	P-PSDB; AAB30567.	
XX		
PT	Novel h19G5 polypeptides capable of regulating signal transduction and	
PT	exhibiting kinase activity useful for identifying antibodies to treat	
PT	cardiac diseases, and additional mediators of signal transduction.	
XX		
XX	Claim 4; Page 57-59; 81pp; English.	
PS		
CC	The present sequence encodes a human protein with putative function in	
CC	signal transduction. The polypeptide is designated H19G5. The protein is	
CC	capable of regulating signal transduction and exhibits kinase activity.	
CC	The H19G5 transcript is expressed in the heart. H19G5 polypeptides and	
CC	polynucleotides are useful for preventing or treating a cardiac disease,	
CC	such as congestive heart failure, dilated congestive cardiomyopathy,	
CC	hyperthrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve	
CC	disease, aortic valve disease or tricuspid valve disease, angina	
CC	pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial	
CC	or renovascular hypertension, arteriosclerosis, atherosclerosis and	
CC	cardiac tumors in humans. The polypeptide is also useful for detecting	
CC	the expression of a protein capable of regulating signal transduction or	
CC	the expression of a protein capable of acting as a donor or acceptor	
CC	molecule of a phosphate group. The monoclonal antibodies can be used as	
CC	probes for detecting discrete antigens expressed by tissue or cell	
CC	samples, and therefore used in humans for localization and monitoring of	
CC	microbial infection	
XX		
SQ	Sequence 4175 BP; 778 A; 1459 C; 1283 G; 655 T; 0 U; 0 Other;	
	Query Match 100.0%; Score 1001; DB 5; Length 4175;	
	Best Local Similarity 100.0%; Pred. No. 1.9e-197;	
	Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 GGCAGTCACTGACACTGGCTGCGAGGTGTACAGCCAGCAGCTGCCAGGCCACCTGG 60	
DB	2578 GCCAGTCACTGACACTGGCTGCCAGGTGTACGCCAGCAGCTGCCAGGCCACCTGG 2637	
QY	61 AGCAAGACGGAGCGCCCTCGGAGAGCAGAGCCGCTGCTCATCTCTGCCACCTCAAG 120	
DB	2638 AGCAAGACGGAGCGCCCTCGGAGAGCAGAGCCGCTGCTCATCTCTGCCACCTCAAG 2697	
QY	121 AACTTCAGCTTTGACATCTCTGGTGGTGGCTGAGGACCTGGGTGTGTACCTGTC 180	
DB	2698 AACTTCAGCTTTGACATCTCTGGTGGTGGCTGAGGACCTGGGTGTGTACCTGTC 2757	
QY	181 AGCGTAGCAATGCGCTGGGAGAGTCACACACCGGCGTCTCTCCGGAGGCGAGCGC 240	
DB	2758 AGCGTAGCAATGCGCTGGGAGAGTCACACACCGGCGTCTCTCCGGAGGCGAGCGC 2817	
QY	241 CCCTCATCTTCGCATCCCGGATATCGGGAGGTGTACGGATGGGGTCTGCTGGTC 300	
DB	2818 CCCTCATCTTCGCATCCCGGATATCGGGAGGTGTACGGATGGGGTCTGCTGGTC 2877	
QY	301 TGGAGCCCGTGGAACTCTACGGCCCTGTGACCTTACATTTGTCAGTGCAGCCCTAGAAGGC 360	
DB	2878 TGGAGCCCGTGGAACTCTACGGCCCTGTGACCTTACATTTGTCAGTGCAGCCCTAGAAGGC 2937	
QY	361 GGCAGCTGGACCACTGGCCCTCCGACATCTTTGACTGCTGCTACCTGACCGACGACCTC 420	
DB	2938 GGCAGCTGGACCACTGGCCCTCCGACATCTTTGACTGCTGCTACCTGACCGACGACCTC 2997	
QY	421 TCCCGGGGTGGACCTTACACTTCCGACCGGCATGTGTACGAAAGGAGGAGAAATGGGTCCC 480	
DB	2998 TCCCGGGGTGGACCTTACACTTCCGACCGGCATGTGTACGAAAGGAGGAGAAATGGGTCCC 3057	
QY	481 TACAGCAGCCCTCGGAGCAAGTCTCTCTCGGAGGGGCCACGCCACCTGGCCTCTGAGGAG 540	

PR 09-NOV-2000; 2000US-0247672P.
PR 16-NOV-2000; 2000US-0249565P.
PR 22-NOV-2000; 2000US-0252730P.
PR 01-DEC-2000; 2000US-0250807P.
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Gururajan R, Baughn MR, Wallia NK, Elliott VS, Xu Y, Arvizu C;
PI Yao MG, Ramkumar J, Ding L, Tang Y, Hafalia AJA, Nguyen DB;
PI Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM, Lal PG;
PI Recipon SA, Lu DAM, Borowsky ML, Thornton M, Swarnaker A;
PI Thangavelu K, Khan FA, Ison CH;
XX
XX WPI; 2002-454603/48.
DR P-PSDB; AAE24151.
DR
XX
XX New human kinase polypeptide, for diagnosing, preventing and treating
PT cancer, immune system disorders, growth and development disorders,
PT cardiovascular disorders and lipid disorders.
PT
XX
PS Claim 5; Page 209-210; 210pp; English.
XX
XX The invention relates human kinases (PKIN) and their corresponding
CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,
CC treating and preventing cancer, an immune system disorder (e.g., acquired
CC immune deficiency syndrome (AIDS), Addison's disease, allergy, asthma,
CC atherosclerosis, multiple sclerosis, psoriasis), disorders affecting
CC growth and development (e.g., arteriosclerosis, cirrhosis, hepatitis),
CC cardiovascular disorder (e.g., hypertension, myocardial infarction,
CC Goodpasture's syndrome), and a lipid disorder (e.g., fatty liver,
CC Gaucher's disease, Niemann-Pick's disease, hypercholesterolemia,
CC hyperlipidaemia, obesity), and for assessing the effects of exogenous
CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a
CC condition or a disease associated with the expression of PKIN in a
CC biological sample. A composition comprising PKIN or an agonist or
CC antagonist of PKIN is useful for treating a disease or condition
CC associated with decreased or increased expression of functional PKIN.
CC PKIN is useful in a number of drug screening techniques and to analyse
CC the proteome of a tissue or cell type. PKIN DNA is useful for creating
CC knockin humanised animals or transgenic animals to model human diseases,
CC and in somatic or germline gene therapy. The present sequence is human
CC PKIN CDNA
XX
XX Sequence 5454 BP; 1019 A; 1853 C; 1686 G; 896 T; 0 U; 0 Other;
SQ

Query Match 100.0%; Score 1001; DB 6; Length 5454;
Best Local Similarity 100.0%; Pred. No. 2e-197;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGCAGTCAGTGACACTGGCTGCCAGGTGTACGCCAGCCAGCAGTGGCCAGCCACTGG 60
Db 3892 GGCAGTCAGTGACACTGGCTGCCAGGTGTACGCCAGCCAGCAGTGGCCAGCCACTGG 3951
Qy 61 AGCAAGACGAGCGCCCTCGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 120
Db 3952 AGCAAGACGAGCGCCCTCGAGAGCAGCAGCGTGTCTCATCTCTGCCACCTCAAG 4011
Qy 121 AACTTCCAGCTTCTGACCATCTCTGGTGTGGCTGAGGACCTGGGTGTGTACACCTGC 180
Db 4012 AACTTCCAGCTTCTGACCATCTCTGGTGTGGCTGAGGACCTGGGTGTGTACACCTGC 4071
Qy 181 AGCGTGAGCAATGGCTGGGACAGTACCAACACGCGGCGTCTCCGGAAGGCGAGGCG 240
Db 4072 AGCGTGAGCAATGGCTGGGACAGTACCAACACGCGGCGTCTCCGGAAGGCGAGGCG 4131
Qy 241 CCCTCATCTTCGCGATATCGGAGGTGTACGGGATGGGGTCTCTGGTC 300
Db 4132 CCCTCATCTTCGCGATATCGGAGGTGTACGGGATGGGGTCTCTGGTC 4191
Qy 301 TGGAGCCCGTGGGAATCTACGGCCCTGTGACCTACATTGTGCAGTGCAGCTAGAGGC 360
Db 4192 TGGAGCCCGTGGGAATCTACGGCCCTGTGACCTACATTGTGCAGTGCAGCTAGAGGC 4251

Qy 361 GGCAGCTGGACCAACACACTGGGCTCCGACATCTTTGACTGCTGCTACCTGACGCAAGCTC 420
Db 4252 GGCAGCTGGACCAACACACTGGGCTCCGACATCTTTGACTGCTGCTACCTGACGCAAGCTC 4311
Qy 421 TCCCGGGGTGGCACTTACACCTTCCGCAAGCAGCATGTGTACCAAGGAGGAGGATGGGTCCC 480
Db 4312 TCCCGGGGTGGCACTTACACCTTCCGCAAGCAGCATGTGTACCAAGGAGGAGGATGGGTCCC 4371
Qy 481 TACAGCAGCCCTCGAGAGCAAGTCTCTCTGGAGGGGCCAGCCACTTGGCTCTTGAGAGAG 540
Db 4372 TACAGCAGCCCTCGAGAGCAAGTCTCTCTGGAGGGGCCAGCCACTTGGCTCTTGAGAGAG 4431
Qy 541 GAGAGCCAGGGGGGTGAGCCCAACCCCTGCCAGCACAAGACCTTCGCATTCAGACACA 600
Db 4432 GAGAGCCAGGGGGGTGAGCCCAACCCCTGCCAGCACAAGACCTTCGCATTCAGACACA 4491
Qy 601 CAGATCCAGAGGGGGCGCTTCAGCGTGTGGCGCAATGCTGGGAGAAAGCCAGCGGGCGG 660
Db 4492 CAGATCCAGAGGGGGCGCTTCAGCGTGTGGCGCAATGCTGGGAGAAAGCCAGCGGGCGG 4551
Qy 661 GCGTGGCGGCCAAGATCATCCCTTACACCCCAAGGACAAAGACGAGTGTCTGCGGAA 720
Db 4552 GCGTGGCGGCCAAGATCATCCCTTACACCCCAAGGACAAAGACGAGTGTCTGCGGAA 4611
Qy 721 TACAGAGCCCTCAAGGGGCTGGCCACCCGACCTGGCCAGCTGCAGCGAGCCCTACCTC 780
Db 4612 TACAGAGCCCTCAAGGGGCTGGCCACCCGACCTGGCCAGCTGCAGCGAGCCCTACCTC 4671
Qy 781 AGCCCCGGGACCTGGTGTCTCATCTGGAGTGTGTCTTGGGCGCGAGCTGTCCCTTCG 840
Db 4672 AGCCCCGGGACCTGGTGTCTCATCTGGAGTGTGTCTTGGGCGCGAGCTGTCCCTTCG 4731
Qy 841 CTGGCCGAGAGGGGCTCTTACTCAGATCCGAGTGAAGGACTACCTGTGCGAGATGTTG 900
Db 4732 CTGGCCGAGAGGGGCTCTTACTCAGATCCGAGTGAAGGACTACCTGTGCGAGATGTTG 4791
Qy 901 AGTCCACCCAGTACCTGTCACACACGACACATCTGACCTGGACCTGAGGTCCGAGAAC 960
Db 4792 AGTCCACCCAGTACCTGTCACACACGACACATCTGACCTGGACCTGAGGTCCGAGAAC 4851
Qy 961 ATGATCATACCGAATACAACTGCTCAAGGTCTGGACCT 1001
Db 4852 ATGATCATACCGAATACAACTGCTCAAGGTCTGGACCT 4892

RESULT 5
ABX11641
ID ABX11641 standard; cDNA; 8106 BP.
XX
AC ABX11641;
XX
DT 09-MAY-2003 (first entry)
XX
DE Human serine/threonine or protein kinase 59079, cDNA.
XX
KW Human; ss; gene; serine/threonine kinase; protein kinase; 59079;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW haemolytic anaemia; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW protein kinase disorder; cellular proliferative disorder; cancer;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 5'UTR 1..71
FT /*tag= a
FT CDS 72..7964
FT /*tag= b
FT /product= "Kinase 59079"
FT /note= "This CDS is specifically claimed in claim 2"

PD	26-OCT-2000.	
XX		
XX	11-APR-2000; 2000WO-US009488.	
XX		
XX	16-APR-1999; 99US-0129553P.	
PR		
XX	(SCIO-) SCIOS INC.	
PA		
XX		
XX	Zeng W, Stanton L, Kong H;	
PI		
XX	WPI: 2001-007013/01.	
DR	P-PSDB; AAB30569.	
DR		
XX		
XX	Novel h19G5 polypeptides capable of regulating signal transduction and	
PT	exhibiting kinase activity useful for identifying antibodies to treat	
PT	cardiac diseases, and additional mediators of signal transduction.	
PT		
XX		
PS	Claim 4; Page 65-68; 81pp; English.	
XX		
XX	The present sequence encodes a splice variant of human in signal	
CC	transduction polypeptide. The polypeptide is designated H19G5. The	
CC	protein is capable of regulating signal transduction and exhibits kinase	
CC	activity. The H19G5 transcript is expressed in the heart. H19G5	
CC	polypeptides and polynucleotides are useful for preventing or treating a	
CC	cardiac disease, such as congestive heart failure, dilated congestive	
CC	cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,	
CC	mitral valve disease, aortic valve disease or tricuspid valve disease,	
CC	angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,	
CC	arterial or thrombovascular hypertension, arteriosclerosis, atherosclerosis	
CC	and cardiac tumours in humans. The polypeptide is also useful for	
CC	detecting the expression of a protein capable of regulating signal	
CC	transduction or the expression of a protein capable of acting as a donor	
CC	or acceptor molecule of a phosphate group. The monoclonal antibodies can	
CC	be used as probes for detecting discrete antigens expressed by tissue or	
CC	cell samples, and therefore used in humans for localization and	
CC	monitoring of microbial infection	
XX		
SQ	Sequence 7928 BP; 1535 A; 2644 C; 2488 G; 1261 T; 0 U; 0 Other;	
	Query Match 99.8%; Score 999.4; DB 5; Length 7928;	
	Best Local Similarity 99.9%; Pred. No. 4.5e-197;	
	Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1 GGCAGTCACTGACACTGGCCCTGCCAGGTGTTCAGCCAGCAGCTGCCAGGCCACCTGG 60	
Db	6372 GGCCAGTCAGTGACACTGGCCTGCCAGGTGTTCAGCCAGCAGCTGCCAGGCCACCTGG 6431	
Qy	61 AGCAAGACGGAGCCCCCTGGAGAGCAGCAGCCGTGTCTCATCTTGCACCCCTCAAG 120	
Db	6432 AGCAAGACGGAGCCCCCTGGAGAGCAGCAGCCGTGTCTCATCTTGCACCCCTCAAG 6491	
Qy	121 AACTTCAGCTTCTGACCATCCTGGTGGTGGCTGAGGACCTGGGGTGTACACCTGC 180	
Db	6492 AACTTCAGCTTCTGACCATCCTGGTGGTGGCTGAGGACCTGGGGTGTACACCTGC 6551	
Qy	181 AGCGTAGCAATGCGCTGGGGACAGTGACACACCGGCGCTCTCCGGAAGGCAGAGCGC 240	
Db	6552 AGCGTAGCAATGCGCTGGGGACAGTGACACACCGGCGCTCTCCGGAAGGCAGAGCGC 6611	
Qy	241 CCCTCATCTTGGCCATGCCCGATATCGGGAGGTGTACCGGATGGGGTGTCTGCTGGTC 300	
Db	6612 CCCTCATCTTGGCCATGCCCGATATCGGGAGGTGTACCGGATGGGGTGTCTGCTGGTC 6671	
Qy	301 TGGNAGCCCGTGGNAATCCTACGGCCCTGTGACCTACATTGTGCACTGCAGCTAGAGGC 360	
Db	6672 TGGNAGCCCGTGGNAATCCTACGGCCCTGTGACCTACATTGTGCACTGCAGCTAGAGGC 6731	
Qy	361 GGCACTGGACCACTAGCTGGCCCTCCGACATCTTTGACTGCTGTACTGTACCCAGCAAGCTC 420	
Db	6732 GGCACTGGACCACTAGCTGGCCCTCCGACATCTTTGACTGCTGTACTGTACCCAGCAAGCTC 6791	
Qy	421 TCCCGGGTGGCACCTTACACTTCCGCACGCGCATGTGTCAAGCAAGGAGGAATGGGTCCC 480	

Db	6792	TCCCGGGTGGCACCTACACCTTTCCGACGGCAGTGTGTACGAAAGGAGGAATGGGTGCC	6951
Qy	481	TACAGCAGCCCCCTCGAGCAAGTCTCTCTGGAGGGGCCACGCCACCTGGGCTCTGAGGAG	540
Db	6852	TACAGCAGCCCCCTCGAGCAAGTCTCTCTGGAGGGGCCACGCCACCTGGGCTCTGAGGAG	6911
Qy	541	GAGAGCCAGGGGGGTTCAGGCCAACCCCTGCCAGCACAAAGACCTTCGCATTCCAGACA	600
Db	6912	GAGAGCCAGGGGGGTTCAGGCCAACCCCTGCCAGCACAAAGACCTTCGCATTCCAGACA	6971
Qy	601	CAGATCCAGAGGGGGCGCTTCACGGTGGTGGCGCAATGCTGGGAGAGGCCAGGGGGCGG	660
Db	6972	CAGATCCAGAGGGGGCGCTTCACGGTGGTGGCGCAATGCTGGGAGAGGCCAGGGGGCGG	7031
Qy	661	GCCTGGCCGCCCAAGATCATCCCTTACCACCCCAAGGACAAAGACAGCAGTGTCTGCGCGAA	720
Db	7032	GCCTGGCCGCCCAAGATCATCCCTTACCACCCCAAGGACAAAGACAGCAGTGTCTGCGCGAA	7091
Qy	721	TACAGAGCCCTCAAGGGCCTTGCGCCACCCGCACTTGGCCAGCTGCACGAGCCTACCTC	780
Db	7092	TACAGAGCCCTCAAGGGCCTTGCGCCACCCGCACTTGGCCAGCTGCACGAGCCTACCTC	7151
Qy	781	AGCCCCGGGACCTGGTGTCTCATCTTGGAGCTGTCTTGGGGCCGAGTGTCTCCCTGC	840
Db	7152	AGCCCCGGGACCTGGTGTCTCATCTTGGAGCTGTCTTGGGGCCGAGTGTCTCCCTGC	7211
Qy	841	CTGGCCGAGAGGCCCTCTACTCAGAATCCGAGGTGAAGGACTACCTGTGCGCAGATGTTG	900
Db	7212	CTGGCCGAGAGGCCCTCTACTCAGAATCCGAGGTGAAGGACTACCTGTGCGCAGATGTTG	7271
Qy	901	AGTGCCACCCAGTACCTGCAACACGACATCCTGCACCTCGACCTGAGGTCGAGAAC	960
Db	7272	AGTGCCACCCAGTACCTGCAACACGACATCCTGCACCTCGACCTGAGGTCGAGAAC	7331
Qy	961	ATGATCATCACCAATACACCTGCTCAAGTCTGTGGACCT	1001
Db	7332	ATGATCATCACCAATACACCTGCTCAAGTCTGTGGACCT	7372
RESULT 11			
ADR41256			
ID	ADR41256 standard; cDNA; 876 BP.		
XX	ADR41256;		
XX	07-OCT-2004 (first entry)		
XX	Human CD-like molecule HAHBE26 cDNA, SEQ ID NO:55.		
DE	Human; CD-like molecule; cluster of differentiation; diagnosis;		
KW	prevention; immune disorder; immunodeficiency; autoimmune disorder;		
KW	blood-related disorder; haematological disorder; haemostatic disorder;		
KW	thrombolytic disorder; hyperproliferative disorder; cancer; tumour;		
KW	apoptotic disorder; cardiovascular disorder; respiratory disorder;		
KW	angiogenic disorder; neovascularisation; neurological disorder;		
KW	endocrine disorder; reproductive system disorder; infectious disease;		
KW	gastrointestinal disorder; drug screening; tissue regeneration;		
KW	chemotaxis; gene therapy; antibody therapy; drug targeting;		
KW	chromosome mapping; forensic analysis; immunophenotyping; cytostatic;		
KW	haemostatic; tranquiliser; vulnery; antinflammatory; nephrotropic;		
KW	cardant; antiallergic; anti-HIV; antirheumatic; antiarthritis;		
KW	antipneumonia; immunosuppressive; vasotropic; neuroprotective;		
KW	antithyroid; thyromimetic; gynaecological; virucide; hepatotropic;		
KW	antibacterial; dermatological; gene; ss.		
OS	Homo sapiens.		
XX	WO200226930-A2.		
XX	04-APR-2002.		
PD	25-SEP-2001; 2001WO-US029838.		
XX			

```
PR 26-SEP-2000; 2000US-0235484P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Birse CE;
XX
XX WPI; 2002-405050/43.
XX P-PSDB; ADR41432.
XX
XX Novel polynucleotides and polypeptides useful for treating, preventing or
XX ameliorating cardiovascular, renal, neurovascular, and autoimmune
XX disorders.
XX
XX Claim 4; SEQ ID NO 55; 1243bp; English.
XX
XX The invention relates to 167 novel human CD (cluster of differentiation)-
XX like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid:11)-
XX
XX Sequence 876 BP; 178 A; 279 C; 268 G; 141 T; 0 U; 10 Other;
XX
XX Query Match 38.5%; Score 385; DB 7; Length 876;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-70;
XX Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 617 GCTTCAGCGTGTGGCGCAATCTGGGAGAGCCAGCGGGCGGCGTGGCCGCAAGA 676
XX Db 125 GCTTCAGCGTGTGGCGCAATCTGGGAGAGCCAGCGGGCGGCGTGGCCGCAAGA 184
XX
XX QY 677 TCATCCCTTACCACCCCAAGGACAGACAGAGTGTGGGATACAGGCGCTCAAGG 736
XX Db 185 TCATCCCTTACCACCCCAAGGACAGACAGAGTGTGGGATACAGGCGCTCAAGG 244
XX
XX QY 737 GCCTGGCCACCCGACCTGGCCCAAGCTGCACGAGCTACTCTAGCCCCCGGACCTGG 796
XX Db 245 GCCTGGCCACCCGACCTGGCCCAAGCTGCACGAGCTACTCTAGCCCCCGGACCTGG 304
XX
XX QY 797 TGCTCATCTTGAGCTGTGCTCTGGGCCCGAGCTGCTCCCTGCTGCGCCAGAGGGCCT 856
XX Db 305 TGCTCATCTTGAGCTGTGCTCTGGGCCCGAGCTGCTCCCTGCTGCGCCAGAGGGCCT 364
XX
XX QY 857 CCTACTCAGATCCGAGGTGAAGGACTACCTGTGCGAGATCTTGAGTGCCACCCAGTACC 916
XX Db 365 CCTACTCAGATCCGAGGTGAAGGACTACCTGTGCGAGATCTTGAGTGCCACCCAGTACC 424
XX
XX QY 917 TGCAACACCCAGCACATCTTCACCTGGACCTGAGGTCCGAGAACATGATCATCACCGAAT 976
XX Db 425 TGCAACACCCAGCACATCTTCACCTGGACCTGAGGTCCGAGAACATGATCATCACCGAAT 484
XX
XX QY 977 ACACCTGCTCAAGTCTGGACCT 1001
XX Db 485 ACACCTGCTCAAGTCTGGACCT 509
XX
XX RESULT 12
XX AAC25654
XX ID AAC25654 standard; cDNA; 182 BP.
XX
XX AC AAC25654;
XX
XX DT 06-OCT-2000 (first entry)
XX
XX DE Human secreted protein 5' EST, SEQ ID NO: 29729.
XX
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX OS Homo sapiens.
XX
XX EN EP1033401-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 21-FEB-2000; 2000EP-00200610.
XX

XX 26-FEB-1999; 99US-0122487P.
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 29729; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX
XX Sequence 182 BP; 34 A; 69 C; 57 G; 19 T; 0 U; 3 Other;
XX
XX Query Match 16.5%; Score 164.8; DB 3; Length 182;
XX Best Local Similarity 97.8%; Pred. No. 1.1e-24;
XX Matches 174; Conservative 3; Mismatches 0; Indels 1; Gaps 1;
XX
XX QY 613 GGCGCGCTTCAGCGTGTGGCGCAATGCTGGGAGAGCCAGCGGGCGGCGCTGGCGGCC 672
XX Db 1 GGCGCGCTTCAGCGTGTGGCGCAATGCTGGGAGAGCCAGCGGGCGGCGCTGGCGGCC 60
XX
XX QY 673 AGATCATATCCCTTACCAACCCCAAGGACAGACAGTGTGCGCGAATACGAGGCCCTC 732
XX Db 61 AGATCATATCCCTTACCAACCCCAAGGACAGACAGTGTGCGCGAATACGAGGCCCTC 120
XX
XX QY 733 AAGGGCTGCGCCACCCGACCTGGCCAGCTGCA-CGACAGCTACTCTAGCCCCCGG 789
XX Db 121 AAGGGCTGCGCCACCCGACCTGGCCAGCTGCGGAGCCGCTACTCTAGCCCCCGG 178
XX
XX RESULT 13
XX ADE47671
XX ID ADE47671 standard; DNA; 9930 BP.
XX
XX AC ADE47671;
XX
XX DT 29-JAN-2004 (first entry)
XX
XX DE Human NOV14a gene SEQ ID NO:33.
XX
XX KW ds; gene; human; cardiant; antiarteriosclerotic; hypotensive;
XX immunosuppressive; dermatological; anorectic; cytostatic; antidiabetic;
XX haemostatic; anti-HIV; antiasthmatic; antibacterial; virucide;
XX neuroprotective; nootropic; antiparkinsonian; antilipaemic; gene therapy;
XX vaccine.
XX
XX OS Homo sapiens.
XX
XX FN WO2003076642-A2.
XX
XX PD 18-SEP-2003.
XX
XX PF 02-AUG-2002; 2002WO-US024459.
XX
```


KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
XX Homo sapiens.

OS WO200073469-A2.

PN PD 07-DEC-2000.

XX PF 26-MAY-2000; 2000WO-US014842.

XX PR 28-MAY-1999; 99US-0136503P.

XX PA (SUGE-) SUGEN INC.

XX PI Plowman GD, Martinez R, Whyte D, Sudersanam S;

XX DR WPI; 2001-032161/04.

XX DR P-PSDB; AAB65635.

XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers.

PS Example 1; Fig 2; 310pp; English.

XX The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure, oxidative-
CC stress related disorders, chronic inflammatory bowel disease, chronic
CC inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis,
CC psoriasis, rhinitis, autoimmunity, diabetes, cancers and reproductive
CC disorders

XX SQ Sequence 7710 BP; 1446 A; 2534 C; 2447 G; 1283 T; 0 U; 0 Other;

Query Match	15.0%	Score 150;	DB 4;	Length 7710;
Best Local Similarity	59.7%	Pred. No. 2e-21;		
Matches 252;	Conservative 0;	Mismatches 170;	Indels 0;	Gaps 0;
QY 580 AAGACCTTTCGATTCACACACAGATCCAGAGGGCGCGCTTCAGCGTGGTGGCGCAATGC	639			
Db 5947 AAACCTTACACCTTCTCGAGGAGAAAGCCAGGGCGCGCTTGGTGGTGGCGAGCGTGC	6006			
QY 640 TGGGAGAGCCAGCGGGCGCGCTGGCCGCAAGATCATCCCTACACCCCAAGGAC	699			
Db 6007 CGGGAGAAATGCCACGGGGCGAACGTTCTGGGCCAAAGATCGTGCCCTATGCTGCCGAGGCG	6066			
QY 700 AAGACGACGTGCTCGCGAATACAGAGCGCTCAAGGGCGCTCGCGCACCCGACCTGGCC	759			
Db 6067 AAGCCGCGGTCTTCAGAGGTACAGAGGTGCTGGCGACCTGACACAGCGGATCATG	6126			
QY 760 CAGCTGCACGACGCTACTCAGCCCCCGGACCTGGTGCTCATCTTGGAGCTGTGCTCT	819			
Db 6127 TCCCTGCACGAGGCGCTACATACCCCTCGGTACCTCGTCTCATTTGCTGAGAGCTGTGGC	6186			
QY 820 GGGCCCGAGCTGCTCCCTGCTGCGGAGAGGGCGCTCTACTCAGATCCGAGGTGAAG	879			
Db 6187 AACCGGGAACCTCTGTGGGCTCAGTGACAGGTTCCGGTATTTCTGAGGATGACGTGGCC	6246			
QY 880 GACTACCTGTGCAGATGTTAGTGCACCCAGCTACCTGCACCAACGACACATCCTGCAC	939			
Db 6247 ACTTACATGGTGACGTGCTACAGGCTTGAGTACCTACCTCCACGGCCACACGTGCTCCAC	6306			

QY 940 CTGACCTGAGTCCGAGAACATGATCATCACCAGATACACCTGCTCAAGGTGCTGCAC	999
Db 6307 CTAGACATCAAGCCAGACAACTGCTGCTGGCCCTGACAATGCCCTCAAGATTGTGCAC	6366
QY 1000 CT 1001	
Db 6367 TT 6368	

Search completed: March 20, 2005, 18:54:57
Job time : 586.044 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:15:10 ; Search time 581.044 Seconds
(without alignments)
10198.307 Million cell updates/sec

Title: US-10-077-130-4_COPY_3120_4120

Perfect score: 1001

Sequence: 1 aaagatgggaagagctgag.....tcagagcccaaggcggtgtt 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:.*
1: geneseqn1980s:.*
2: geneseqn1990s:.*
3: geneseqn2000s:.*
4: geneseqn2001as:.*
5: geneseqn2001bs:.*
6: geneseqn2002as:.*
7: geneseqn2002bs:.*
8: geneseqn2003as:.*
9: geneseqn2003bs:.*
10: geneseqn2003cs:.*
11: geneseqn2003ds:.*
12: geneseqn2004as:.*
13: geneseqn2004bs:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	24120	8 ABX11642	Abx11642 Human ser
2	965	96.4	20489	12 ADQ22881	Adq22881 Human sof
3	248.4	24.8	2768	8 ABX71198	Abx71198 Novel hum
4	218.6	21.8	1251	10 ADC30509	Adc30509 Human nov
5	218.6	21.8	2488	4 AAH18453	Aah18453 Human cdn
6	218	21.8	3956	13 ADR08183	Adr08183 Full leng
7	140.2	14.0	2737	4 AA159251	Aa159251 Human pol
8	140.2	14.0	5378	13 ADR67197	Adr67197 Human bla
9	140.2	14.0	5382	10 ADD14722	Add14722 Human src
10	130.2	13.0	1458	12 ADL24314	Adl24314 AW755252-
11	129.8	13.0	3999	4 AA161037	Aa161037 Human pol
12	123.2	12.3	1645	5 Aaf24162	Aaf24162 Human sec
13	123.2	12.3	1948	2 ABV74346	Abv74346 Human il-
14	123.2	12.3	1949	2 ABV74347	Abv74347 Human il-
15	123.2	12.3	2170	6 ABQ54970	Abq54970 Human ova
16	123	12.3	2184	4 AAH18256	Aah18256 Human cdn
17	107	10.7	2559	8 ACC46215	Acc46215 Human dit
18	104	10.4	724	4 AAH08421	Aah08421 Human cdn
19	97.4	9.7	14061	6 ABV99363	Abv99363 Human nov
20	97.4	9.7	14109	6 ABV99362	Abv99362 Human nov

21	95.8	9.6	2534	11 ADM03261	Adm03261 Human cdn
22	95.8	9.6	7564	8 ABZ24581	Abz24581 Human cel
23	88.4	8.8	2254	3 AAC93403	Aac93403 Human sec
24	88.4	8.8	2254	5 AAF24183	Aaf24183 Human sec
25	71	7.1	572	5 AAS42508	Aas42508 Human cdn
26	68	6.8	716	6 ABK35726	Abk35726 cDNA sequ
27	66.8	6.7	9591	4 AAS28906	Aas28906 Human inm
28	66.8	6.7	9591	4 AAL03355	Aal03355 Human rep
29	66.8	6.7	9591	10 ADB31747	Adb31747 Human nov
30	66.8	6.7	12415	4 AAS28907	Aas28907 Human inm
31	66.8	6.7	12415	4 AAL03356	Aal03356 Human rep
32	66.8	6.7	12415	10 ADB31748	Adb31748 Human nov
33	66.8	6.7	20565	4 AAK89083	Aak89083 Human dig
34	66.8	6.7	20565	4 AAS28908	Aas28908 Human inm
35	66.8	6.7	20565	4 AAL03357	Aal03357 Human rep
36	66.8	6.7	20565	10 ADB31749	Adb31749 Human nov
37	62.2	6.2	30191	12 ADQ97654	Adq97654 Mouse can
38	62	6.2	31595	10 ADF81661	Adf81661 Leukaemia
39	62	6.2	81940	4 AAS05390	Aas05390 Human tit
40	62	6.2	81940	6 ABK64829	Abk64829 Human ben
41	62	6.2	81940	12 ADQ17315	Adq17315 Human sof
42	62	6.2	93801	9 ABX13540	Abx13540 Human RGS
43	62	6.2	103052	13 ADQ89963	Adq89963 Antagonis
44	61.6	6.2	1266	6 ABK35725	Abk35725 cDNA sequ
45	61.6	6.2	20448	4 ABL19989	Ab119989 Drosophil

ALIGNMENTS

RESULT 1

ABX11642
ID ABX11642 standard; cDNA; 24120 BP.

AC ABX11642;

DT 09-MAY-2003 (first entry)

DE Human serine/threonine or protein kinase 12599, cDNA.

KW Human; ss; Gene; serine/threonine kinase; protein kinase; 12599;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.

OS Homo sapiens.

PH Key Location/Qualifiers

FT 5'UTR 1..71

FT CDS /tag= a

FT /tag= b

FT /product= "Kinase 12599"

FT /note= "This CDS is specifically claimed in claim 2"

FT 3'UTR 23979..24120

FT /tag= c

PN US2002168742-A1.

XX 14-NOV-2002.

XX 15-FEB-2002; 2002US-00077130.

XX 15-FEB-2001; 2001US-0269201P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Kapeller-Libermann R, Acton SL;

XX


```
PT sarcoma.
PS
XX
XX Example 2; SEQ ID NO 5701; 210pp; English.
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytosstatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 20489 BP; 3917 A; 5815 C; 6983 G; 3225 T; 0 U; 549 Other;

Query Match 96.4%; Score 965; DB 12; Length 20489;
Best Local Similarity 96.4%; Pred. No. 6.9e-218;
Matches 965; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 AAAGATGGGAGAGAGCTGAGCTCCAGCTTGAAGTGCATGTAGAGCCAAAGGCTGCAGA 60
DB 3120 AAAGATGGGAGAGAGCTGAGCTCCAGCTTGAAGTGCATGTAGAGCCAAAGGCTGCAGA 3179

QY 61 CGGAGGCTGGTGTGCAGCAGCAGGCAAGACGGATGCCGGGACTACAGCTGCAGGCCC 120
DB 3180 CGGAGGCTGGTGTGCAGCAGCAGGCAAGACGGATGCCGGGACTACAGCTGCAGGCCC 3239

QY 121 AGGGGCCAGAGGGTCTCTTCGCTGCACATCACAGAGCCCAAGATGATGTTTCAAG 180
DB 3240 AGGGGCCAGAGGGTCTCTTCGCTGCACATCACAGAGCCCAAGATGATGTTTCAAG 3299

QY 181 GAGCAGTCAGTGCATTAATGAGTGCAGCTGAGCGGGGGCCAGTGCCATGCTGAGCTGT 240
DB 3300 GAGCAGTCAGTGCATTAATGAGTGCAGCGGNNNNNNNNNNNNNNNNNNNNNNNNNN 3359

QY 241 GAGGTGGCCCGCCAGCCAGAGGTGACGTGTGTAAGAGGTGGGAAGAGCTGAGCTCC 300
DB 3360 NNNNNGGCCCGCCAGCCAGAGGTGACGTGTGTAAGAGGTGGGAAGAGCTGAGCTCC 3419

QY 301 AGCTCAAAAGTGGGATGAGAGTCAAAAGGTGCAACGGAGGCTGGTGTGCAACAGGCG 360
DB 3420 AGCTCAAAAGTGGGATGAGAGTCAAAAGGTGCAACGGAGGCTGGTGTGCAACAGGCG 3479

QY 361 GGCAAGCAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGCCAGAGAGTCTCTTCCAC 420
DB 3480 GGCAAGCAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGCCAGAGAGTCTCTTCCAC 3539

QY 421 CTGCATCATCACAGAGCCCAAGGGGGTGTGCGAAGAGGAGCAGTCAGTGCATAATGAGTGTG 480
DB 3540 CTGCATCATCACAGAGCCCAAGGGGGTGTGCGAAGAGGAGCAGTCAGTGCATAATGAGTGTG 3599

QY 481 CAGGCTGAGGGGGGACCACTGCCATGCTGAGCTGTGAGGCTGGGGCCAGAGAGTCTCTTCCAC 540
DB 3600 CAGGCTGAGGGGGGACCACTGCCATGCTGAGCTGTGAGGCTGGGGCCAGAGAGTCTCTTCCAC 3659

QY 541 GTGAGCTGTTACAGGAGGGGAGAGTGCAGCTCCAGCTCAAAAGTACGATGAGGTC 600
DB 3660 GTGAGCTGTTACAGGAGGGGAGAGTGCAGCTCCAGCTCAAAAGTACGATGAGGTC 3719

QY 601 AAGGCTGCACAGAGGCTGGTGTGTCAGCAGAGTGGGCAAGCAGATGCTGGGAGGTAC 660
DB 3720 AAGGCTGCACAGAGGCTGGTGTGTCAGCAGAGTGGGCAAGCAGATGCTGGGAGGTAC 3779

QY 661 AGCTCGAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCATCACAGAGCCCAAGGCA 720
DB 3780 AGCTCGAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCATCACAGAGCCCAAGGCA 3839

QY 721 GTGTTTCCAGGAGCAGTGTGTCATTAATGAGGTGCGGACTGAGGCGAGGGGCCAGTGCC 780
```

```
Db 3840 GTGTTTCCAGGAGCAGTGTGTCATTAATGAGGTGCGGACTGAGGCGAGGGGCCAGTGCC 3899
QY 781 ACATGAGCTGTGAGTGGGCCAGGCCCAGACAGAGTGCAGTGTGTAACAGATGGGAAG 840
Db 3900 ACATGAGCTGTGAGTGGGCCCAGGCCCAGACAGAGTGCAGTGTGTAACAGATGGGAAG 3959
QY 841 AAGCTGAGCTCCAGTTCGAAAGTGGCATAGAGCTCGGGCTGCATCGGCAGCTGGTG 900
Db 3960 AAGCTGAGCTCCAGTTCGAAAGTGGCATAGAGCTCGGGCTGCATCGGCAGCTGGTG 4019
QY 901 GTGCAGCAGGCGAGCCAGGAGATGCTGGGAGTACACCTGTGAGGCTGGGGGCCAGCGG 960
Db 4020 GTGCAGCAGGCGAGCCAGGAGATGCTGGGAGTACACCTGTGAGGCTGGGGGCCAGCGG 4079
QY 961 CTCTCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGT 1001
Db 4080 CTCTCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGT 4120

RESULT 3
ABX71198
ID ABX71198 standard; cDNA; 2768 BP.
XX
XX AC ABX71198;
XX
XX DT 05-MAR-2003 (first entry)
XX
XX Novel human cDNA sequence #423.
Human; gene; se; nervous system disorder; peripheral neuropathy;
Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
Crohn's disease; anaphylaxis; proliferation; chemotactic;
differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
haemostatic; antiinflammatory; expressed sequence tag; EST.
Homo sapiens.
WO200281731-A2.
17-OCT-2002.
29-JAN-2002; 2002WO-US001222.
30-JAN-2001; 2001US-00774528.
(HYSE-) HYSEQ INC.
(GOOD/) GOODRICH R W.
Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
WPI; 2003-058563/05.
Novel polypeptide useful for treating neurodegenerative diseases, myeloid
or lymphoid cell disorders, bone disorders, mechanical and traumatic
disorders, coagulation disorders, and inflammatory diseases.
Claim 1; Page; 612pp; English.
This invention relates to the cDNA sequences encoding an isolated novel
human polypeptide. The protein encoded by the nucleic acid of the
invention is useful for treating central and peripheral nervous system
diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
Alzheimer's disease); autoimmune disease (e.g. systemic lupus
erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
```

CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia).
CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
CC bacterial, viral or fungal infections; allergic conditions such as
CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
CC cancer and tumours; and inflammatory diseases (e.g. Crohn's
CC disease, anaphylaxis). The protein may be used to inhibit the growth,
CC infection or function of infectious agents such as bacteria, fungi,
CC viruses, or to effect bodily characteristics, biorhythms or circadian
CC cycles of rhythms. The protein may also have
CC proliferation/differentiation, stem cell growth factor, haematopoiesis
CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory,
CC activities. The cDNA sequences of the invention are useful for expressing
CC recombinant protein for analysis. The present sequence represents a novel
CC human cDNA sequence of the invention, this sequence is an expressed
CC sequence tag (EST) and was identified using subtractive hybridisation
xx

QY	786	GAGCTGTGAGGTGGCCAGGCCCAGACAGAGAGGTGACGTGGTGTACAAAGATGGGGAAGACT	845
Db	2591	GAGCTGTGAGGTGGCCAGGACCAGATGGAGGTGACGTGGTACAAAGACGGGAAGAGTT	2650
QY	846	GAGCTCCAGTTGCGAAGTGCATAGAGGCTGCGGGCTGCATGCGGAGCTGTGGTGGTGC	905
Db	2651	GAGCTCCAGCTCGAAGTGCACGTGGAGGCTGTGGGCTGTATCGGAGGCTGGTGGTGC	2710
QY	906	GCAGCGAGGCCAGGCAGATGCTGGGGAGTACACCTGTGAGGCTGGGGG	953
Db	2711	GCAGGTGGCCAGGCAGACTCCGATAGTACAGCTGTGAAGCCAGGTG	2758
RESULT 4			
ADCS0509			
ID	ADCS0509	standard; cDNA; 1251 BP.	
XX	AC	ADCS0509;	
XX	DT	18-DEC-2003 (first entry)	
XX	DE	Human novel cDNA sequence, SEQ ID NO:591.	
XX	KW	Human; diagnostic; drug screening; forensics; gene mapping;	
KW	KW	biodiversity assessment; Parkinson's disease; Alzheimer's disease;	
KW	KW	neurodegenerative diseases; anaemia; platelet disorder; wound; burns;	
KW	KW	ulcers; osteoporosis; autoimmune disease; cancer;	
KW	KW	molecular weight marker; food supplement; antiparkinsonian; nootropic;	
KW	KW	neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary;	
KW	KW	ant ulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;	
KW	KW	gene therapy; chromosome 17; gene; ss.	
XX	OS	Homc saplens.	
XX	PN	WO2003029271-A2.	
XX	PD	10-APR-2003.	
XX	PF	24-SEP-2002; 2002WO-US030474.	
XX	PR	24-SEP-2001; 2001US-0324631P.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	PI	Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;	
PI	PI	Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;	
PI	PI	Haley-Vicente D, Drmanac RT;	
XX	XX	WPI; 2003-371981/35.	
DR	DR	P-FSDB; ADC31480.	
XX	XX		
PT	PT	New polynucleotide and polypeptide useful for diagnosing, preventing or	
PT	PT	treating conditions such as neurodegenerative diseases, anemias, platelet	
PT	PT	disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or	
PT	PT	cancer.	
XX	XX	Claim 1; SEQ ID NO 591; 1185pp; English.	
PS	PS		
XX	XX		
CC	CC	The invention relates to 971 novel human cDNA sequences (ADC29919-	
CC	CC	ADC30889) and the polypeptides they encode (ADC30890-ADC31960). The	
CC	CC	invention also relates to nucleic acid sequences over 99% identical with	
CC	CC	the novel human cDNAs. The invention additionally encompasses expression	
CC	CC	vectors and host cells comprising a nucleic acid of the invention; the	
CC	CC	recombinant production of a polypeptide of the invention; an antibody	
CC	CC	against a polypeptide of the invention; a method of detecting	
CC	CC	polynucleotides or polypeptides of the invention; and methods of	
CC	CC	identifying a compound which binds to a polypeptide of the invention. The	
CC	CC	invention further discloses methods of preventing, treating or	
CC	CC	ameliorating a medical condition; kits comprising polynucleotide probes	
CC	CC	and/or monoclonal antibodies for carrying out the methods of the	
CC	CC	invention; methods for the identification of compounds that modulate the	
CC	CC	expression or activity of the polynucleotide and/or polypeptide; and 767	
CC	CC	contig sequences corresponding to the cDNA sequences of the invention	

CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a specifically
 CC claimed human cDNA sequence of the invention. Note: the sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1251 BP; 329 A; 293 C; 363 G; 266 T; 0 U; 0 Other;

Query Match 21.8%; Score 218.6; DB 10; Length 1251;
 Best Local Similarity 86.1%; Pred. No. 8.2e-42;
 Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 705 CACAGAGCCCAAGGAGTGTTCACAGGAGGAGTGTGATATAGTGGGACTGA 764
 DB 891 CCCAGAGCCCAAGGAGTGTTCACAGGAGGAGTGTGATATAGTGGGACTGA 950
 QY 765 GGCAGGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCACAGAGGAGTGCAGTG 824
 DB 951 GCGGGGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCACAGAGGAGTGCAGTG 1010
 QY 825 GTACAAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGCACATAGAGGCTGCGGGCTG 884
 DB 1011 GTACAAGGATGGGAAGAGCTGAGTTCAGCTCGAAAGTGCAGGCGGTGGGGCTG 1070
 QY 885 CATCGGCGAGCTGGTGTGACAGGAGGAGGAGGAGGAGTGTGAGTGCAGTGTGA 944
 DB 1071 CACAGGAGGCTGGTGTGACAGGAGGAGGAGGAGGAGTGTGAGTGCAGTGTGA 1130
 QY 945 GCGTGGGGGCCAGGCGGCTCTCTTCCAGCTGGATGTTTCAG 985
 DB 1131 GGCAGGGGGTTCAGCAGCTCTCTTCCGCTGCAGGTGCGAG 1171

RESULT 5
 ID AAH18453
 XX AAH18453 standard; cDNA; 2488 BP.
 AC AAH18453;
 XX
 XX 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:18549.
 XX
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 XX EP1074617-A2.
 XX
 XX 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 95JP-00248036.
 PR 27-AUG-1999; 95JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.

XX Claim 8; SEQ ID NO 18549; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH24446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

SQ Sequence 2488 BP; 561 A; 637 C; 716 G; 574 T; 0 U; 0 Other;

Query Match 21.8%; Score 218.6; DB 4; Length 2488;
 Best Local Similarity 86.1%; Pred. No. 9.7e-42;
 Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 705 CACAGAGCCCAAGGAGTGTTCACAGGAGGAGTGTGATATAGTGGGACTGA 764
 DB 891 CCCAGAGCCCAAGGAGTGTTCACAGGAGGAGGAGTGTGATATAGTGGGACTGA 950
 QY 765 GGCAGGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCACAGAGGAGTGCAGTG 824
 DB 951 GCGGGGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCACAGAGGAGTGCAGTG 1010
 QY 825 GTACAAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGCACATAGAGGCTGCGGGCTG 884
 DB 1011 GTACAAGGATGGGAAGAGCTGAGTTCAGCTCGAAAGTGCAGGCGGTGGGGCTG 1070
 QY 885 CATCGGCGAGCTGGTGTGACAGGAGGAGGAGGAGGAGTGTGAGTGCAGTGTGA 944
 DB 1071 CACAGGAGGCTGGTGTGACAGGAGGAGGAGGAGGAGTGTGAGTGCAGTGTGA 1130
 QY 945 GCGTGGGGGCCAGGCGGCTCTCTTCCAGCTGGATGTTTCAG 985
 DB 1131 GGCAGGGGGTTCAGCAGCTCTCTTCCGCTGCAGGTGCGAG 1171

RESULT 6
 ID ADR08183/c
 ID ADR08183 standard; cDNA; 3956 BP.
 XX
 AC ADR08183;
 XX
 XX 04-NOV-2004 (first entry)
 XX
 DE Full length human cDNA useful for treating neurological disease Seq 1689.
 XX
 KW gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;

osteoporosis; neurological disease; Alzheimer's disease;
Parkinson's disease; dementia; short memory; cancer;
sense or motor function; emotional reaction; fear response; panic;
osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
tranquilliser.
XX Homo sapiens.
XX EP1447413-A2.
XX 18-AUG-2004.
XX 12-FEB-2004; 2004EP-00003145.
XX 14-FEB-2003; 2003JP-00102207.
XX 09-MAY-2003; 2003JP-00131452.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
XX Wakamatsu A, Ishii S, Nagai K, Irie R;
XX WPI: 2004-583265/57.
XX P-PSDB; ADR10139.
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
XX Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX Claim 1; SEQ ID NO 1689; 2686pp; English.
XX This invention relates to novel, isolated full length human cDNA
XX molecules and the encoded proteins thereof. Specifically, it refers to
XX cDNA clones obtained by an oligo-capping method, where none of these
XX clones are identical to any known human mRNAs. The present invention
XX describes an immunosay to identify agonists and antagonists, as well as
XX antibodies, antisense molecules and siRNAs that can all be used to bind
XX to and modulate expression of the cDNA molecules. As such, these
XX molecules are useful for diagnostic markers or therapeutic targets for
XX the various diseases or morbid states. In particular, they are useful in
XX gene therapy for treating osteoporosis, neurological disease, Alzheimer's
XX disease, Parkinson's disease, dementia, short memory and various cancers,
XX as well as for maintaining equilibrium of sense or motor function, and
XX for treating emotional reaction, fear response and panic. Accordingly,
XX they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
XX cytostatic and tranquilliser activities. This polynucleotide is a full
XX length human cDNA sequence of the invention. NOTE: This sequence is not
XX given in the sequence listing of the specification but can be obtained on
XX CD-ROM from the European Patent Office, Vienna Sub-office.
XX Sequence 3956 BP; 882 A; 1211 C; 933 G; 930 T; 0 U; 0 Other;
SQ Query Match 21.8%; Score 218; DB 13; Length 3956;
Best Local Similarity 84.5%; Pred. No. 1.5e-41;
Matches 245; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 144 CTTGACATCATCAGAGCCCAAGATGATTTTGAAGAGAGCAGTCAGTCGATATGAGGT 203
DB 1141 CTTGTCCATCCAGAGCCCAAGAGTGTGTTTGAAGAGAGCAGTCAGTCGAGGT 1082
QY 204 GCAGGCTCAGGCGGGCCAGTCGATGCTGAGTGTGAGTGCAGCCAGCCAGACGGA 263
DB 1081 GCAGCGGAGCGGGGGCCAGTCGATGCTGAGTGTGAGTGCAGCCAGCCAGATGGA 1022
QY 264 GGTGACGTGGTACAGGATGGGAAGAGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGGT 323
DB 1021 GGTGACATGGTACAGGAGCGGAGAGAGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGG 962
QY 324 CAAGAGGTGCACACGAGGCTGGTGTCCACACGCGGGCCAAAGCAGATGCTGGGGAGTA 383
DB 961 CAGCGGCTACACACGAGGCTGGTGTGAGTGCAGCAGCGGGCCAGCGGATGCTGGGGAGTA 902
QY 384 CAGCTGTGAGCTGGGGGCCAGAGAGTCTCTCCACTGCACATCACAG 433
|||||

Db 901 CAGCTGTGAGCGGGGGGCCAACAGCTGTCTCTCCGCTGCACGTGGCAG 852
RESULT 7
AAI59251
ID AAI59251 standard; cDNA; 2737 BP.
XX AC AAI59251;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 1454.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US034263.
XX 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00552317.
XX 20-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00682191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX (HVSE-) HVSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX P-PSDB; AAM40095.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX Claim 1; SEQ ID NO 1454; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
SQ Sequence 2737 BP; 484 A; 840 C; 899 G; 514 T; 0 U; 0 Other;
Query Match 14.0%; Score 140.2; DB 4; Length 2737;
Best Local Similarity 48.4%; Pred. No. 3.3e-23;
Matches 484; Conservative 0; Mismatches 508; Indels 9; Gaps 3;

Db 1996 GAGGTGGACGAGAGAGACCCCTCTGTGCTTGGTACAAGGACGGCAGAGGTGAGGAG 2055
QY 301 AGCTCAAAAGTGGGATGAGGTCAAAAGGTTGCACACGAGGCTGGTGTGCACACAGCG 360
Db 2056 AGTGACTTGTGTGCTGGAGATGAGGGGCCCATCGCCGCTGGTGTGCTGCCGCCACC 2115
QY 361 GGAAGAAGCAGATGCTGGGGAGTACAGCTGTGTAGGTGTGGGGCCAGAGAGTCTCTTTCAC 420
Db 2116 CAGCCCTCAGACGGGGGAGTTTCACTGCTGCTGGAGATGAGTGTGCTTACTTACT 2175
QY 421 CTGCATACACAGCCCAAGGGGTGTTTGAAGGAGCAGTCACTGATCAATAGAGTG 480
Db 2176 GTACCATACAGAGCTCTCTCTGTGATCG---TGTATCCAGCGGCAAGGTGTATGT 2232
QY 481 CAGGCTGAGGGCGGACCACTGCCATGCTGAGCTGTGAGTGGCCAGCCACAGCGAG 540
Db 2233 CGAGCGTGGCGCTGGAGCGTGTGTGCTGCTGAGCTATGCGGCCCTTGGCAGAG 2292
QY 541 GTGAGTGTGTACAAGGACGGGAAGGTGAGTCTCAGCTCAAAAGTACGATGAGGTC 600
Db 2293 GTGCGCTGGACCAAGGATGGAGAGGTGTGTGAGAGCCCGCGCTGCTCTGCAGAAG 2352
QY 601 AAGGCTGCACACGAAGCTGTGTGCTGAGAGTGGGCAAGCAGATGCTGGGAGTAC 660
Db 2353 GAAGACACTGTCCCGCGCTGTGTGCTGCTGCGCTGTCCAGCTCGAGGACTCCGGGAGTAC 2412
QY 661 AGCTCGAGGTGGGGGCGCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGCA 720
Db 2413 TTGTGTGAATGTGAGATGAGTGGGCTCTCTTCACTGTACCGTACAGAACCC---CCA 2469
QY 721 GTGTTTGCAGGAGCAGTGTGTGTGATATGAGTGTGGGACTGAGGCGAGGGCCAGTGCC 780
Db 2470 GTGCGGATCATATACCTCGCATGAGGTGACCTTGTATGCTGCGCTGACCTTGGAGTGTG 2529
QY 781 ACACGTAGCTGTGAGGTGGCCAGCCAGCAGAGAGTACGTGTGTACAAGATGGGAG 840
Db 2530 GTGCTGATGTGTAATCTGTCTCGGAGGATGCCCCGTGTGCTGTGAGAGGATGGCCACGCTGCGCTGTG 2589
QY 841 AAGCTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTGGGGCTGTGATCGCGCAGCTGTG 900
Db 2590 GAAGTGGAGGAGAGCGAGGCCCTGTGTGTGAGAGGATGGCCACGCTGCGCTGTG 2649
QY 901 GTGCAGAGGAGCGCCAGGCGAGTCTGGGAGGTACACTGTGAGGCTGGGGGCGCAGCG 960
Db 2650 CTACCTGTCTCAGCCCGAGGACGGGGCGAGTTTGTATGTGATGCTGTGAGATCACTCG 2709
QY 961 CTCTCTTCCACCTGATGTTTCAGAGCCCAAGCGGTGTT 1001
Db 2710 GCCTTCTTCACTGTCACTGTACAGAGCCCTCCAGTGCAGTT 2750

RESULT 9
ID ADD14722 standard; cDNA; 5382 BP.
XX AC ADD14722;
XX 01-JAN-2004 (first entry)
XX Human src biomarker polynucleotide SEQ ID NO:116.
XX predictor set; protein tyrosine kinase activity modulator;
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
KW gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
XX Homo sapiens.
XX WO2003062395-A2.
XX 31-JUL-2003.
XX 17-JAN-2003; 2003WO-US001981.

PR 18-JAN-2002; 2002US-0350061P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX Huang F, Fairchild CR, Lee FY, Shaw P;
XX WPI; 2003-636735/60.
DR P-PSDB; ADD14123.
XX New polynucleotides and polypeptides for predicting the activity of
PT compounds that interact with protein tyrosine kinases and/or protein
PT tyrosine kinase pathways.
XX Claim 2; SEQ ID NO 116; 139pp; English.
XX The present invention describes a predictor set comprising a plurality of
CC polynucleotides or polypeptides whose expression pattern is predictive of
CC the response of cells to treatment with a compound that modulates protein
CC tyrosine kinase activity or members of the protein tyrosine kinase
CC pathway. Also described: (1) predicting whether a compound is capable of
CC modulating the activity of cells, comprising obtaining a sample of cells,
CC determining whether the cells express a plurality of markers, and
CC correlating the expression of the markers to the compound's ability to
CC modulate the activity of the cells; (2) a plurality of cell lines for
CC identifying polynucleotides and polypeptides whose expression levels
CC correlate with compound sensitivity or resistance of cells associated
CC with a disease state; and (3) identifying polynucleotides and
CC polypeptides that predict compound sensitivity or resistance of cells
CC associated with a disease state, comprising subjecting the plurality of
CC cell lines to one or more compounds, analysing the expression pattern of
CC a microarray of polynucleotides or polypeptides, and selecting
CC polynucleotides or polypeptides that predict the sensitivity or
CC resistance of cells associated with a disease state by using the
CC expression pattern of the microarray. The polynucleotides and
CC polypeptides have cytostatic activities, and can be used in gene therapy.
CC The polynucleotides and polypeptides are useful in predicting the
CC activity of compounds that interact with protein tyrosine kinases and/or
CC protein tyrosine kinase pathways. These may be used in determining drug
CC sensitivity in patients to allow the development of individualized
CC genetic profiles which aid in treating diseases and disorders (e.g.
CC cancer) based on patient response at a molecular level. The present
CC sequence is used in the exemplification of the present invention.
XX SQ Sequence 5382 BP; 943 A; 1699 C; 1798 G; 942 T; 0 U; 0 Other;
Query Match 14.0%; Score 140.2; DB 10; Length 5382;
Best Local Similarity 48.4%; Pred. No. 3.9e-23;
Matches 484; Conservative 0; Mismatches 508; Indels 9; Gaps 3;
QY 1 AAAGATGGGAAGAGCTGAGCTCCAGCTTGAAAGTGCATGTAGAGGCAAGGCTGCAGA 60
Db 1760 AAGGATGGGCAGAAAGGTGGAGGAGAGCGAGTTGCTGTGTGAAGATGGATGGCGCAA 1819
QY 61 CGGAGGCTGTGTGTGACGAGGAGGAGCAGACGAGATGCCGGGACTACAGCTGGAGGCC 120
Db 1820 CACCGTCTGATCTCGCTGAGGGCCAAAGTCCAGGACAGTGGCGAGTTTGATGTGAGGACA 1879
QY 121 AGGGGCCAGAGGGTCTCTTCCGCTGCACATCACAGAGCCCAAGATGATGTTTCAAG 180
Db 1880 GAAGGGGTCTCGGCCCTTCTTCGGCGTCACTGTCCAAGATCTCCCGTGCACATCG---TG 1936
QY 181 GAGCAGTCACTGATTAATAGGTGCAAGGCTGAGCGGGGCCAGTGCATCTGAGCTGT 240
Db 1937 GACCCCGAGAACATGTTTGTGTCATGCCATAAATCCAGTGTGTGTCATGCTGGCCTGT 1996
QY 241 GAGGTGGCCAGGCCAGCAGAGGAGTGCAGTGTGTACAGATGGGAGAGCTGAGTCC 300
Db 1997 GAGGTGACCGAGAGGACGCCCCCTGTGCTGTGTGTACAGGACGGGAGGAGGAG 2056
QY 301 AGCTCAAAAGTGGGCATGAGAGGTCAAGAGGTGCACACGAGGCTGGTGTGCTCCACAGCG 360
Db 2057 AGTGACTTGTGTGTGTGGAGATGAGGGGCCCATGCGCGCTGTGTGCTGCCGCCACC 2116

QY 361 GCGAAGCAGATGCTGGGGAGTACAGCTGTAGGCTGGGGCCAGAGAGTCTCTTCCAC 420
Db |||||
QY 2117 CAGCCCTCAGACGGGGGAGTTTCAGTGGCTGGAGATGAGTGGCTTACT 2176
Db |||||
QY 421 CTGCACATCAGACGCCAAGGGGGTGTTCGAAGGAGCAGTCAAGTCAATAAGAGTG 480
Db |||||
QY 2177 GTCACCATCAGATGTCTCTCGTGATCG--TGATCCAGCGCAAGGTGTATGTG 2233
QY 481 CAGGCTAGGGGGGACCACCTGCCATGCTAGCTGTAGGTGGCCAGCCCGCAGCGAG 540
Db |||||
QY 2234 GCAGCAGTGGCGCTGGAGCGTGTGTGCTGACCTGTAGCTATGCGGCGCTGGCGAG 2293
QY 541 GTGAGCTGTACAGGAGCGGAAGAGCTGAGCTCCAGTCAAAAGTACGATGAGGTC 600
Db |||||
QY 2294 GTGCGCTGGACCAAGATGGAGAGAGGTGTGGAGAGCCCGCGCTGCTCTCGAGAAG 2353
QY 601 AAGGCTGCACACGAAGGCTGTGTAGTCAGCAGAGGTGGCAAGCAGATGCTGGGAGTAC 660
Db |||||
QY 2354 GAAGACACTGTCCGCGCTGTGTGCTGCCGCTGTCCAGCTCGAGACTCCGCGAGTAC 2413
QY 661 AGCTCGAGGCTGGGGCCAGAGAGTCTCTTTTCAATGTCACATCAGAGAGCCCAAGGCA 720
Db |||||
QY 2414 TTGTGTGAATTTGACGATGAGTGGCGCTCTTCACTGTCAAGTCAAGAGAGCC--CCA 2470
QY 721 GTGTTTCCAGGAGCAGTGTGTGATGATGAGGTGCGGACTGAGGCGAGGGCCAGTCC 780
Db |||||
QY 2471 GTGCGGATCATATACCTCGGATGAGGTGACCTTGTATGCGCGTACCTTGGAGTGTG 2530
QY 781 ACACGTAGCTGTAGGTGGCCAGGCCAGACAGAGGTGACGTGTGTAACAAGGATGGGAG 840
Db |||||
QY 2531 GTGCTGATGTGTAATGCTCTCGGAGGATGCCCTGTGGCTGTATCAAGGATGGGCTG 2590
QY 841 AAGCTGAGCTCAGTTCGAAGTGGCATAGAGGCTGCGGGTGTGATCGGCGAGTGTG 900
Db |||||
QY 2591 GAAAGTGGAGGAGAGCGAGCGCTGTGTGCTGGAGAGGATGGGCCACGCTGCGCGCTGTG 2650
QY 901 GTGCAGCAGGAGGCGAGCGATGCTGGGAGTACCTGTGAGGCTGGGGCCAGCGG 960
Db |||||
QY 2651 CTACCTGTGCTCAGCCGAGAGCGGGGGGAGTTTGTATGTGATGCTGGAGATGACTCG 2710
QY 961 CTCTCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTT 1001
Db |||||
QY 2711 GCCTTCTTCACTGTCTACAGAGCTCCAGTGCAGTT 2751

RESULT 10

ID ADL24314
XX ADL24314 standard; DNA; 1458 BP.
XX AC ADL24314;
XX DT
XX 03-JUN-2004 (first entry)
XX DE AW755252-interacting protein mPN34854 coding sequence SEQ ID NO: 13.
XX KW AW755252; cardiant; gene therapy; ischaemic heart disease;
KW myocardial infarction; cardiac failure; dilated cardiomyopathy;
KW angina pectoris; hypertrophia cordis; mPN34854; db; gene.
XX OS Unidentified.
XX PN WO2004019880-A2.
XX PD 11-MAR-2004.
XX PF 26-AUG-2003; 2003WO-US026997.
XX PR 28-AUG-2002; 2002US-0406613P.
XX PR 18-SEP-2002; 2002US-0411657P.
XX PA (TAKE) TAKEDA CHEM IND LTD.
XX PI Hensel C, Sakamoto T;

XX WPI; 2004-239111/22.
XX New protein complex, useful for preparing a composition for treating diseases e.g., ischemic heart disease, myocardial infarction, cardiac failure, dilated cardiomyopathy, hypertrophia cordis or angina pectoris.
XX Claim 129; Fig 7; 151pp; English.
XX The present invention provides a new isolated protein complex comprising a first protein comprising AW755252 or its homologue or derivative or a fragment interacting with a second protein consisting of MFHL2, mPN34854, mPRHL, mCTEX-1, ACTN2, mACTN4 and mMRJ, or a homologue or derivative or fragment. The protein complex is useful for preparing a composition for treating diseases e.g., ischaemic heart disease, myocardial infarction, cardiac failure, dilated cardiomyopathy, hypertrophia cordis or angina pectoris. The present sequence is the coding sequence of the mPN34854 protein, which interacts with the AW755252 protein.
XX Sequence 1458 BP; 294 A; 390 C; 470 G; 304 T; 0 U; 0 Other;

Query Match 13.0%; Score 130.2; DB 12; Length 1458;
Best Local Similarity 47.7%; Pred. No. 6.6e-21;
Matches 477; Conservative 0; Mismatches 518; Indels 6; Gaps 3;

QY 1 AAAGATGGGAAGAGCTGAGCTCCAGCTTGAAGTGCATGTAGAGCCAAAGGCTGCAGA 60
Db |||||
QY 352 AAGGATGGCAGGAGGTGGAGAGAGTGCATCATCTGATTAGAAAATAAAGGCGCCCAT 411
QY 61 CGGAGGCTGTGTGTCAGCAGCGAGCAAGACGATGCCGGGACTACAGCTGCAGGCC 120
Db |||||
QY 412 CACCGCTGTGTGTCAGCTGAGCGCGGCCCTCCGAGGGGGGAGTTCAGTGTGCGCA 471
QY 121 AGGGGCCAGAGGGTCTCTCCGCTGCACATCAGAGCCCAAGATGATGTTTCAAG 180
Db |||||
QY 472 GGAGATGAACGTGCTACTTCAAGTTACCATCACAGATGCTTCTCGTGGATGCTTA- 530
QY 181 GAGCAGTCAGTGCATATGAGTGCAGCTGAGGGGGGGGCGAGTGCCTGCTGAGCTGT 240
Db |||||
QY 531 -CCCCAGTAGCAAGTGCATGTGGCAGCCATACGCCCTAGAGC-GTGTGCTGCTGACCTGT 588
QY 241 GAGTGGCCAGGCGCCAGACGCGAGTGCAGTGGTGTACAGAGATGGGAAGCTGAGCTCC 300
Db |||||
QY 589 GAGTGTGCCACCTCGGGCTGAGTGGCTGGACCAAGATGGGGAGAGGTAGTGAG 648
QY 301 AGCTCAAAAGTGGGATGGAGGTCAAAGGGTGCACAGAGGGTGGTGTCTGCCACAGCG 360
Db |||||
QY 649 AGCCAGCACTGCTCTCGGAGAGAGACACCATCCGCCCTGGTGTCTGCCCTCTGTC 708
QY 361 GGCAAGCAGATGCTGGGGAGTACAGTGTGAGGCTGGGGGGCCAGAGAGTCTCTTCCAC 420
Db |||||
QY 709 CAGCTTGAGGATTTCTGGCGAGTACCTGTGTGAAATCCATGATGATGCTGCGCTTCCCTTACC 768
QY 421 CTGCACATCAGAGAGCCCAAGGGGTGTTTCCGAGGAGCAGTCACTGCATATGAGGTG 480
Db |||||
QY 769 ATCAGCTCAGAGAGCCCTGTGGGATCATATACCCCGAGAGAGGTGA---CCTTA 825
QY 481 CAGGCTGAGGCGGGAGCCACCTGCTGAGCTGTGAGGTGGCCCGCAGAGCGAG 540
Db |||||
QY 826 CACGCGTGTGAGTTTGGATGTGTGTGCTCACCTGTGAGTGTCTAGAGAGGATGCTCCT 885
QY 541 GTGAGTGTGTAAGGAGCGGAAGAGTGCAGTCCAGCTCAAAAGTACGATGAGGAGTGC 600
Db |||||
QY 886 GTACGCTGTACAAGGATGGGTTAGAGGTGGAGGAGTGAAGCCCTGCTGCTCCAGAGC 945
QY 601 AAGGCTGCACACGAAGGCTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGGAGTAC 660
Db |||||
QY 946 GATGGCCCTGCTGCCGCTGTGTGCTGCTGCTGCCAGCAGAGAGCGGGGCGAGTTT 1005
QY 661 AGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTTCAACTGTCATCAGAGAGCCCAAGGCA 720
Db |||||
QY 1006 GTGTGTGATGCTGGGGATGATTACGCTTCTTCACTGTCTGCTCACAGCTCCACAGAA 1065

PN WO9918990-A1.
XX 22-APR-1999.
XX 09-OCT-1998; 98WO-US021347.
XX 10-OCT-1997; 97US-0061720P.
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX Young PR;
XX WPI; 1999-277442/23.
XX P-PSDB; ABB98343.
XX Novel interleukin 1 receptor related protein 3 (IL-1RRP3) or variants
XX for, e.g. treatment of Alzheimer's disease.
XX Claim 2; Page 30-31; 43pp; English.
XX The invention relates to polynucleotides encoding interleukin 1 receptor
XX related protein 3 (IL-1RRP3) or variants. The polynucleotides are used to
XX diagnose a disease or susceptibility to a disease in a subject related to
XX expression or activity of the IL-1RRP3 to screen for compounds which
XX stimulate or inhibit the function of IL-1RRP3 or to treat a subject in
XX need of enhanced or inhibited IL1RRP3 activity. Specifically mentioned in
XX the specification is the use of the IL-3RRP3 polynucleotide or
XX polypeptide for treatment of chronic and acute inflammation, arthritis,
XX septicemia, autoimmune disease, transplant rejection, graft-versus-host
XX disease, infection, stroke, ischaemia, acute respiratory disease
XX syndrome, restenosis, brain injury, AIDS, bone diseases e.g.
XX osteoporosis, cancer, atherosclerosis and Alzheimer's disease. The
XX present sequence is that of a human IL-1RRP3 encoding cDNA of the
XX invention. The IL-1RRP3 gene is located on chromosome 2q33
XX
XX Sequence 1948 BP; 397 A; 561 C; 600 G; 386 T; 0 U; 4 Other;
Query Match 12.3%; Score 123.2; DB 2; Length 1948;
Best Local Similarity 49.5%; Pred. No. 3.2e-19;
Matches 376; Conservative 0; Mismatches 378; Indels 6; Gaps 2;
QY 230 TGCTGAGCTGTGAGTGGCCAGCCAGCGAGGTGACGTGTACAGGATGGGAGA 289
Db 825 TGCTGACTGTGAGCTCTCAAGGGTGGACTTCCCGGCAACCTGGTACAAAGATGGCGAGA 884
QY 290 AGCTGAGCTCCAGCTCAAAAGTGGCGATGGAGGTCAAAAGGTGCACACGGAGGTGGTGC 349
Db 885 AGGTGGAGGAGCGAGTTGCTGTGTGTAAGATGGATGGCGCAACACCGTCTGATCC 944
QY 350 TGCCACAGCGCGCAAGCAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGCCAGAGAG 409
Db 945 TGCTGAGGCCAAAGTCCAGGACAGTGGCGAGTTGAGTGCAGGACAGAAAGGGTCTCGG 1004
QY 410 TCTCTCTCCACTGCACATCAGAGCCCAAGGGGTGTTTGGCAAGAGCAGTCACTGTC 469
Db 1005 CCTTCTTCGGGGTCACTGTCCAAAGATCTCCCGTGCACATCG---TGGACCCCGAGAAC 1061
QY 470 ATAATGAGGTGCGAGCTGAGCGGGGACCACTGCCATGCTGAGTGTGAGGTGGCCGAGC 529
Db 1062 ATGTGTTGTCATGCCATAAATTCCGAGTGTGTCATGCTGCTGTGAGGTGGACCGAG 1121
QY 530 CCCAGACGGAGGTGACGTGGTACAGGACGGGAAGAAGCTGAGCTCCAGCTCAAAAAGTAC 589
Db 1122 AGGACGCCCTGTGCGTTGGTACAAAGGACGGGAGGAGTGGAGAGTCACTTCGTGG 1181
QY 590 GCATGGAGGTCAAGGCTGCACAGAGGCTGGTAGTGCACAGAGTGGGCAAGCAGATG 649
Db 1182 TGCTGGGAATGAGGGGGCCCCATCGCCGCTGGTGTCTCCCGCCACCCAGCCCTCAGACG 1241
QY 650 CTGGGGAGTACAGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAG 709
Db 1242 GGGCGGAGTTTCAGTGGCTGCTGGAGATGATGTGCTCTTCACTGTCACTACATCACAG 1301

QY 710 AGCCCAAGGCGAGTCTTTTGCACAGGAGCAGTGTGGTCATAATGAGGTGCGGACTGAGGCAG 769
Db 1302 A---CGTCTCTCTCGTGTGATCCAGCGCAAGGTGTATGTGGCAGCCGCTGCGCC 1358
QY 770 GGGCCAGTGCACACTGAGCTGTGAGGTGGCCAGGCCAGAGGTGACGTGGTACA 829
Db 1359 TGGAGCGTGTGGTGTGACCTGTGAGCTATGCCCGCCCTGGGCGAGGTGGCTGGACCA 1418
QY 830 AGGATGGGAAGAAGCTGAGCTCCAGTTTCGAAAGTGGCGCATAGAGGCTGCGGGCTGCATGC 889
Db 1419 AGGATGGAGAGGAGGTGGTGGAGAGCCCGCGCTGCTCTCCAGAAGAGACACTGTCC 1478
QY 890 GGCAGCTGGTGTGACAGCAGCAGCCAGGAGATGCTGGGGAGTACACCTGTGAGGCTG 949
Db 1479 GCGCCTGGTGTGCTCCCGCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAAATTG 1538
QY 950 GGGCCAGCGGCTCTCTTCCACCTGATGTTTCAGAGCC 989
Db 1539 ACGATGAGTGGCGCTCTTCACTGTCCCGCTCACAGAGTC 1578
RESULT 14
ABV74347
ID ABV74347 standard; cDNA; 1949 BP.
XX
XX AC ABV74347;
XX
XX 21-FEB-2003 (first entry)
XX Human IL-1RRP3 encoding cDNA SEQ ID NO 3.
XX
XX Human; interleukin 1 receptor related protein 3; IL-1RRP3; inflammation;
XX arthritis; septicemia; autoimmune disease; transplant rejection;
XX infection; stroke; AIDS; osteoporosis; cancer; Alzheimer's disease;
XX anti-inflammatory; antiarthritic; antibacterial; immunosuppressive;
XX cerebroprotective; vasotropic; antiviral; osteopathic; cytostatic;
XX antiarteriosclerotic; nootropic; neuroprotective; chromosome 2q33; gene;
XX ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 414..1700
XX FT /*tag= a
XX FT /product= "IL-1RRP3"
XX
XX WO9918990-A1.
XX
XX 22-APR-1999.
XX
XX 09-OCT-1998; 98WO-US021347.
XX
XX 10-OCT-1997; 97US-0061720P.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
XX Young PR;
XX
XX WPI; 1999-277442/23.
XX P-PSDB; ABB98344.
XX
XX Novel interleukin 1 receptor related protein 3 (IL-1RRP3) or variants
XX for, e.g. treatment of Alzheimer's disease.
XX
XX Claim 13; Page 31-32; 43pp; English.
XX
XX The invention relates to polynucleotides encoding interleukin 1 receptor
XX related protein 3 (IL-1RRP3) or variants. The polynucleotides are used to
XX diagnose a disease or susceptibility to a disease in a subject related to
XX expression or activity of the IL-1RRP3 to screen for compounds which
XX stimulate or inhibit the function of IL-1RRP3 or to treat a subject in
XX need of enhanced or inhibited IL1RRP3 activity. Specifically mentioned in
XX the specification is the use of the IL-3RRP3 polynucleotide or

Matches 376; Conservative 0; Mismatches 378; Indels 6; Gaps 2;	
QY 230	TGCTGAGCTGTGAGTGGCCAGAGCCAGACGAGGTGACGTGGTGTACAAAGGATGGGAAGA 289
Db	
1082	TGCTGACTTGTGAGCTCTCAAGGTGGACTTCCCGGCAACCTGGTACAAAGATGGGCAGA 1141
QY	290 AGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGGTCAAAGGGTGCACACGGAGGCTGGTGCC 349
Db	
1142	AGGTGGAGGAGAGCGAGTGTGCTGGTGAAGATGGATGGGCGCAACACACCGTCTGTATCC 1201
QY	350 TGCCACAGCGGGCAAAAGCATGCTGGGGAGTACAGCTGTGAGCTGGGGCCACAGAGAG 409
Db	
1202	TGCTTGAGGCCAAAGTCCAGGACAGTGGCGAGTTTGAAGTGAAGGACAGAAAGGGTCTCGG 1261
QY	410 TCTCTTCCACCTGCACATCACAGAGCCCAAGGGGTGTTTCCGAAGGAGCAGTCAGTGCC 469
Db	
1262	CCTTCTTCGGGTCACTGTCCAAGATCTCCCGTGCACATCG---TGGACCCCGAGAAC 1318
QY	470 ATAATGAGGTGAGGCTGAGCGGGGACCACTGCCATGCTGAGCTGTGAGGTGGCCAGC 529
Db	
1319	ATGTGTCGTGCATGCCATACTTCCGAGTGTGTCACTGGCCTGTGAGGTGGACCGAG 1378
QY	530 CCCACAGGAGTACGCTGCTACAGGACGGGAGAGAGCTGAGCTCCAGCTCAAAAGTAC 589
Db	
1379	AGGACGCCCTGTGCTTGGTACAAAGACGGCGAGAGGTGAGGAGAGTGACTTCGTGG 1438
QY	590 GCATGGAGGTCAAGGGCTGCACACGAAGGCTGGTAGTGCAGCAGGTGGGCGAAAGCAGATG 649
Db	
1439	TGCTGGAGAATGAGGGGCCCCATCGCCGCTGGTGTCTGCCGCCACCCAGGCCCTCAGACG 1498
QY	650 CTGGGGAGTACAGCTGCGAGGCTGGGGGCCAGAGAGTCTCCTTTCAACTGCACATCACAG 709
Db	
1499	GGGGCGAGTTTCAGTGCCTGCTGGAGATGAGTGTGCCCTACTTCACCTGCACCATCACAG 1558
QY	710 AGCCCAAGGCAGTGTTCGCAAGGAGCAGTTGGTGCATAATGAGTGGCGGACTGAGGCAG 769
Db	
1559	A---CGTCTCTCTGGGATCGTGATATCCAGCGGCAAGGTGTATGTGGCAGCCGTGGGCC 1615
QY	770 GGGCCAGTGCCACACTGAGCTGTGAGGTGGGCCCCAGGCCCCAGACAGAGGTGACGTGGTACA 829
Db	
1616	TGGAGCGTGTGGTGTGACCTGTGAGCTATGCCGCCCTGGGCAGAGGTGCGCTGGACCA 1675
QY	830 AGGATGGGAAGAAGCTGAGCTCCAGTTCGAAAGTGGCGCATAGAGGCTCGGGCTGCATGC 889
Db	
1676	AGGATGGAGAGAGGAGTGGTGGAGAGCCCGCGCTGCTCTGCAGAAAGGAGACACTGTCC 1735
QY	890 GGCAGCTGGTGTGCAGCAGCAGCGCCAGGCGAGATGCTGGGGAGTACACCTGTGAGGCTG 949
Db	
1736	GCCGCTGGTGTGCTGCCGCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGAAATTG 1795
QY	950 GGGGCCAGCGCTCTCCTTCCACCTGGATGTTTCAGAGCC 989
Db	
1796	ACGATGAGTCGGCCTCCTTCACTGTACCGTCAACAGTGC 1835

Search completed: March 20, 2005, 18:54:48
Job time : 588.044 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 13:15:10 ; Search time 581.044 Seconds
(without alignments)
10198.307 Million cell updates/sec

Title: US-10-077-130-4_COPY_10500_11500
Perfect score: 1001
Sequence: 1 gccctgtgagtgaggaa.....ttcctagaagatgtgaaaaa 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	20489	12 ADQ22881	Adq22881 Human sof
2	1001	100.0	24120	8 ABX11642	Abx11642 Human ser
3	997.8	99.7	7564	8 ABZ24581	Abz24581 Human cel
4	997.8	99.7	14061	6 ABV99363	Abv99363 Human NOV
5	997.8	99.7	14109	6 ABV99362	Abv99362 Human NOV
6	343.4	34.3	352	5 AAF64470	Aaf64470 Novel hum
7	265	26.5	642	6 ABQ61169	Abq61169 Obscurin
8	265	26.5	707	6 ABQ61180	Abq61180 Obscurin
9	238.8	23.9	2155	6 ABK99965	Abk99965 DNA encod
10	200.6	20.0	898	6 ABL58139	AbL58139 Human pho
11	107.4	10.7	2534	11 ADM03261	Adm03261 Human CDN
12	86	8.6	1005	6 ABN21414	Abn21414 Human ORF
13	62	6.2	3935	11 ADM01849	Adm01849 Human CDN
14	60	6.0	60	6 ABN40768	Abn40768 Human epl
15	59	5.9	93801	9 ABX13540	Abx13540 Human RGS
16	59	5.9	103052	13 ADQ89963	Adq89963 Antagonis
17	58.8	5.9	3999	4 AAI61037	Aai61037 Human pol
18	57.8	5.8	2737	4 AAI59251	Aai59251 Human pol
19	57.8	5.8	5378	13 ADR67197	Adr67197 Human bla
20	57.4	5.7	31595	10 ADF81661	Adf81661 Leukaemia

21	57.4	5.7	81940	4 AAS05390	Aas05390 Human tit
22	57.4	5.7	81940	6 ABK64829	Abk64829 Human ben
23	57.4	5.7	81940	12 ADQ17315	Adq17315 Human sof
24	56.4	5.6	5382	10 ADD14722	Add14722 Human arc
25	53.6	5.4	1948	2 ABV74346	Abv74346 Human IL-
26	53.6	5.4	1949	2 ABV74347	Abv74347 Human IL-
27	53.6	5.4	2170	6 ABQ54970	Abq54970 Human ova
28	52	5.2	1645	5 AAF24162	Aaf24162 Human sec
c 29	50.8	5.1	1337	2 AAZ17263	Aaz17263 Human gen
30	50.6	5.1	1218	3 AAA02488	Aaa02488 Human col
31	50.4	5.0	2000	8 ADA71938	Ada71938 Rice gene
32	49.8	5.0	2254	3 AAC93403	Aac93403 Human sec
33	49.8	5.0	2254	5 AAF24183	Aaf24183 Human sec
34	48.6	4.9	10732	3 AAA10594	Aaa10594 Gene enco
35	47.6	4.8	716	6 ABK35726	Abk35726 CDNA sequ
36	44	4.4	1435	8 ABV74143	Abv74143 Human dua
37	43.8	4.4	1000	3 AAA02484	Aaa02484 Human col
38	43.6	4.4	5452	10 ADC86736	Adc86736 Human GPC
39	43	4.3	7328	4 AAK89081	Aak89081 Human dig
40	43	4.3	20565	4 AAK89083	Aak89083 Human dig
c 41	43	4.3	20565	4 AAS28908	Aas28908 Human inm
42	43	4.3	20565	4 AAL03357	Aal03357 Human rep
c 43	43	4.3	20565	10 ADB31749	Adb31749 Human nov
44	42.4	4.2	1593	3 AAA02504	Aaa02504 Human col
45	42.4	4.2	1945	13 ADR96750	Adr96750 Human dua

ALIGNMENTS

RESULT 1
ADQ22881
ID ADQ22881 standard; DNA; 20489 BP.
XX
AC ADQ22881;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5701.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
(PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 5701; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has

CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 20489 BP; 3917 A; 5815 C; 6983 G; 3225 T; 0 U; 549 Other;
Query Match 100.0%; Score 1001; DB 12; Length 20489;
Best Local Similarity 100.0%; Pred. No. 3.9e-231;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCTCTGGAGTGGAGAGAGGGCCCCGAGAACCTTCAGAGATGGGGACAGATACATCTCTG 60
Db |
QY 10500 GCCCTCTGGAGTGGAGAGAGGGCCCCGAGAACCTTCAGAGATGGGGACAGATACATCTCTG 10559
Db |
QY 61 AGGCAGAGAGGGACACAGCTGTGAGCTGCAGATCTGTGGCTGGCCATGGCGAGCCCGGG 120
Db |
QY 10560 AGGCAGAGAGGGACACAGCTGTGAGCTGCAGATCTGTGGCTGGCCATGGCGAGCCCGGG 10619
QY 121 GAGTACTTGTGTGTGGCGGACAGAGAGGACCTCAGCCAGCTCACCATCAGGGCTCTG 180
Db |
QY 10620 GAGTACTTGTGTGTGGCGGACAGAGAGGACCTCAGCCAGCTCACCATCAGGGCTCTG 10679
QY 181 CCTGCAGGTTTCATAGAGATGTGAAAAACAGAGAGGCCAGAGAGGGGCCACCGCTGTG 240
Db |
QY 10680 CCTGCAGGTTTCATAGAGATGTGAAAAACAGAGAGGCCAGAGAGGGGCCACCGCTGTG 10739
QY 241 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGGAGTGGAGAAAGGGTCTGACACCTC 300
Db |
QY 10740 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGGAGTGGAGAAAGGGTCTGACACCTC 10799
QY 301 AGAGATGGGGACAGATACAGCTCAGGCGAGCAGCGGACTAAATGTGAGCTGCAGATCGT 360
Db |
QY 10800 AGAGATGGGGACAGATACAGCTCAGGCGAGCAGCGGACTAAATGTGAGCTGCAGATCGT 10859
QY 361 GGCCTGGCCATGGCAGACACTTGGGAGTACTCTGCGTGTGCGGCGAGGAGAGACCTCG 420
Db |
QY 10860 GGCCTGGCCATGGCAGACACTTGGGAGTACTCTGCGTGTGCGGCGAGGAGACCTCG 10919
QY 421 GCTATGCTCACCGTCAAGGCTCAACCATCAAGTTTCAGAGGGTCTGAGGAACGAGAG 480
Db |
QY 10920 GCTATGCTCACCGTCAAGGCTCAACCATCAAGTTTCAGAGGGTCTGAGGAACGAGAG 10979
QY 481 GCCACAGAGGGGCAACAGCCGCTGTCGCGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG 540
Db |
QY 10980 GCCACAGAGGGGCAACAGCCGCTGTCGCGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG 11039
QY 541 TGGTGGAGGGGCATGAGACCTCAGAGATGGAGACACACAGCCTGAGGACGAGCGG 600
Db |
QY 11040 TGGTGGAGGGGCATGAGACCTCAGAGATGGAGACACACAGCCTGAGGACGAGCGG 11099
QY 601 GCCAGGTGTGAGCTGCAGATCCGCGGCTCTGTCAGAGGACGCTGGGGAGTACCTGTGC 660
Db |
QY 11100 GCCAGGTGTGAGCTGCAGATCCGCGGCTCTGTCAGAGGACGCTGGGGAGTACCTGTGC 11159
QY 661 ATGTGCGGGAGGAGAGGACCTCAGCCATGCTCACCGTCAAGGCGCATTCCTTCAAGTTC 720
Db |
QY 11160 ATGTGCGGGAGGAGAGGACCTCAGCCATGCTCACCGTCAAGGCGCATTCCTTCAAGTTC 11219
QY 721 ATAGAGGCTCTGAGGATGAAGAGGCCACAGAGGGGACACAGCCACGCTGTGTGTGAG 780
Db |
QY 11220 ATAGAGGCTCTGAGGATGAAGAGGGCCACAGAGGGGACACAGCCACGCTGTGTGTGAG 11279
QY 781 CTGAGCAAGGGCGGACCGGTGGAGTGGAGGAAGGGGACATGAGACCTCAGAGATGGGGAC 840
Db |
QY 11280 CTGAGCAAGGGCGGACCGGTGGAGTGGAGGAAGGGGACATGAGACCTCAGAGATGGGGAC 11339
QY 841 AGACACAGCTTGAGGACGAGCGGTCCAGGTGTGAGCTGCAGATCCGTGGCCCTGGCTGTG 900
Db |
QY 11340 AGACACAGCTTGAGGACGAGCGGTCCAGGTGTGAGCTGCAGATCCGTGGCCCTGGCTGTG 11399

QY 901 GTGGATCCGGGGAGTACTCTGTGTGTGGCGGAGAGAGACCTCAGCCACACTCACT 960
Db |
QY 11400 GTGGATCCGGGGAGTACTCTGTGTGTGGCGGAGAGAGACCTCAGCCACACTCACT 11459
Db |
QY 961 GTCAAGGCGCCCTGCCAGATTCATAGAGATGTGAAAA 1001
Db |
QY 11460 GTCAGGCGCCCTGCCAGATTCATAGAGATGTGAAAA 11500
RESULT 2
ABX11642
ID ABX11642 standard; cDNA; 24120 BP.
XX
XX AC ABX11642;
XX 09-MAY-2003 (first entry)
XX Human serine/threonine or protein kinase 12599, cDNA.
DE Human; ss; gene; serine/threonine kinase; protein kinase; 12599;
XX Cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 5'UTR 1..71 /*tag= a
FT CDS 72..23978 /*tag= b
FT /*tag= b
FT /product= "Kinase 12599"
FT /note= "This CDS is specifically claimed in claim 2"
FT 3'UTR 23979..24120 /*tag= c
XX
XX US2002168742-A1.
XX
XX 14-NOV-2002.
XX
XX 15-FEB-2002; 2002US-00077130.
XX
XX 15-FEB-2001; 2001US-0269201P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Kapeller-Libermann R, Acton SL;
PI WPI; 2003-298729/29.
PI P-PSDB; ABG76187.
DR
DR Novel isolated human protein kinase, designated 59079 or 12599
XX polypeptide, useful as diagnostic and therapeutic agents for preventing
XX cardiovascular diseases, proliferative disorders, and protein kinase
XX disorders.
PS Claim 2; Page 58-84; 119pp; English.
XX
XX The invention relates to an isolated human serine/threonine or protein
XX kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
XX comprising at least 85% identity to the nucleic acids appearing as
XX ABX11641 and ABX11642 or their complement, a naturally occurring variant
XX of the kinases or their fragments. Also included are a non-human host
XX cell containing the nucleic acids, an antibody specific for the proteins,
XX identifying a compound which binds to the kinase (by contacting the
XX kinase or a cell expressing the kinase with a test compound and
XX determining whether the kinase binds to the test compound) and
XX the activity of kinase using the identified compound. The kinases and
XX their encoding nucleic acids are useful as diagnostic and therapeutic

agents for preventing a disease or condition associated with an aberrant or unwanted 59079 or 12599 activity in a subject, including cardiovascular diseases such as heart failure, and myocardial infarction; disorders involving blood vessels such as atherosclerosis, and Kaposi's sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia, Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders such as cancer; and protein kinase disorders such as autoimmune disorders, diabetes mellitus, psoriasis, inflammatory bowel disease, rheumatoid arthritis, and multiple sclerosis (many examples of diseases and disorders are included in the specification). The kinases, their encoding nucleic acids and antibodies are useful in screening assays, detection assays (e.g. forensic biology), and predictive medicine (e.g. diagnostic assays, prognostic assays, and monitoring clinical trials and pharmacogenomics). The kinases and their encoding nucleic acids are useful as query sequences to perform a search against public databases to identify other family members or related sequences. The present sequence encodes the kinase 12599

SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;

Query Match 100.0%; Score 1001; DB 8; Length 24120;
 Best Local Similarity 100.0%; Pred. No. 4.1e-231;
 Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCTGTGGAGTGGAGGAGGCGCCGAGAACCTCAGAGATGGGACAGATACATCTCG 60
 DB 10500 GCCCTGTGGAGTGGAGGAGGCGCCGAGAACCTCAGAGATGGGACAGATACATCTCG 10559

QY 61 AGGCAGAGGGGACACAGCTGTGAGCTGCAGATCTGTGGCCCTGGCCATGGCGAGCGCCGG 120
 DB 10560 AGGCAGAGGGGACACAGCTGTGAGCTGCAGATCTGTGGCCCTGGCCATGGCGAGCGCCGG 10619

QY 121 GAGTACTTGTGTGTGGCGGAGAGAGACCTCAGCACGCTCACCATCAGGGCTCTG 180
 DB 10620 GAGTACTTGTGTGTGGCGGAGAGAGACCTCAGCACGCTCACCATCAGGGCTCTG 10679

QY 181 CCTGCCAGTTCTAGAGAGTGTGAAGAACAGAGGCGCCAGAGAGGGCCACCGCTGTG 240
 DB 10680 CCTGCCAGTTCTAGAGAGTGTGAAGAACAGAGGCGCCAGAGAGGGCCACCGCTGTG 10739

QY 241 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGGAGTGGAGAAAGGGGTCTGAGACCTC 300
 DB 10740 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGGAGTGGAGAAAGGGGTCTGAGACCTC 10799

QY 301 AGAGATGGGACAGATACAGCTCAGGAGGACCGGACTAAATGTGAGCTCAGATTCTG 360
 DB 10800 AGAGATGGGACAGATACAGCTCAGGAGGACCGGACTAAATGTGAGCTCAGATTCTG 10859

QY 361 GGCCTGGCCATGGCAGACACTGGGAGTACTCTGCTGCTGGCGGAGGAGGAGGACCTCG 420
 DB 10860 GGCCTGGCCATGGCAGACACTGGGAGTACTCTGCTGCTGGCGGAGGAGGAGGACCTCG 10919

QY 421 GCTATGCTCACCGTTCAGGGCTCTACCCATCAAGTTTCAGAGGGGTCTGAGGAACGAGAG 480
 DB 10920 GCTATGCTCACCGTTCAGGGCTCTACCCATCAAGTTTCAGAGGGGTCTGAGGAACGAGAG 10979

QY 481 GCCACAGAGGGGCAACCGCTGTGCGGTGTGAGCTGAGCAAGATGCCCGCTGGAG 540
 DB 10980 GCCACAGAGGGGCAACCGCTGTGCGGTGTGAGCTGAGCAAGATGCCCGCTGGAG 11039

QY 541 TGGTGGAGGGGCATGAGACCTCAGAGATGGAGACAGACAGCCTGAGGACGAGCGG 600
 DB 11040 TGGTGGAGGGGCATGAGACCTCAGAGATGGAGACAGACAGCCTGAGGACGAGCGG 11099

QY 601 GCCAGGTGTGAGCTGCAGATCCGGGCTCTGTCGAGAGGACGCTGGGGAGTACTCTGTC 660
 DB 11100 GCCAGGTGTGAGCTGCAGATCCGGGCTCTGTCGAGAGGACGCTGGGGAGTACTCTGTC 11159

QY 661 ATGTGCGGGAAGGAGGAGGACCTCAGCCATGTCTCAGCGTCAGGGCCATGCTTCCAAAGTTC 720
 DB 11160 ATGTGCGGGAAGGAGGAGGACCTCAGCCATGTCTCAGCGTCAGGGCCATGCTTCCAAAGTTC 11219

QY 721 ATAGAGGCTCTGAGGAATGAAGAGGCCACAGAGGGGGACAGCGGCCACGCTGTGTGTGAG 780

Db 11220 ATAGAGGGTCTGAGGATGAGAGGCCACAGAGGGGACACGGCCACGCTGTGTGTGAG 11279
 QY 781 CTGAGCAAGCGCGCACCGGTGGAGTGGAGGAAGGGGATGAGACCTCAGAGATGGGGAC 840
 Db 11280 CTGAGCAAGCGCGCACCGGTGGAGTGGAGGAAGGGGATGAGACCTCAGAGATGGGGAC 11339
 QY 841 AGACACAGCTGAGGACGAGGAGCGGTCCAGGTGTGAGCTGCAGATCCGCTGGCCTGGCTGTG 900
 Db 11340 AGACACAGCTGAGGACGAGGAGCGGTCCAGGTGTGAGCTGCAGATCCGCTGGCCTGGCTGTG 11399
 QY 901 GTGATCCCGGGAGTACTCTGTGTGTGTGGCGGAGAGAGGACCTCAGCCACACTCACT 960
 Db 11400 GTGATCCCGGGAGTACTCTGTGTGTGTGGCGGAGAGAGGACCTCAGCCACACTCACT 11459
 QY 961 GTGATCCCGGGAGTACTCTGTGTGTGTGGCGGAGAGAGGACCTCAGCCACACTCACT 1001
 Db 11460 GTGATCCCGGGAGTACTCTGTGTGTGTGGCGGAGAGAGGACCTCAGCCACACTCACT 11500

RESULT 3
 ABZ24581

ID ABZ24581 standard; cDNA; 7564 BP.

XX AC ABZ24581;

DT 31-MAR-2003 (first entry)

XX Human cell adhesion and extracellular matrix protein 4 cDNA.

XX Cell adhesion and extracellular matrix protein 4; CADECM-4; human;
 anti-HIV; virucide; antiallergic; antiinflammatory; antianaemic;
 antiparkinsonian; nootropic; anticonvulsant; antifertility;
 antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
 cyostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;
 angiot; thymimetic; neuroprotective; osteopathic; antiarthritic;
 antiparasitic; antelmintic; antipsoriatic; uropathic; ophthalmological;
 antineumatic; haemostatic; antibacterial; protozoacide; fungicide;
 gynaecological; titin; gene therapy; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 CDS 241..7227

FT /*tag= a

FT /product= "Human CADECM-4"

PN WO200288322-A2.

XX 07-NOV-2002.

XX 01-MAY-2002; 2002WO-US013874.

XX 02-MAY-2001; 2001US-0288290P.

XX 21-MAY-2001; 2001US-0292468P.

XX 15-JUN-2001; 2001US-0298616P.

XX 28-JUN-2001; 2001US-0301672P.

XX 04-JAN-2002; 2002US-0345008P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding L;

PI Hillman J, Baughn MR, Kallick DA, Lee S, Warren BA, Xu Y, Tran UK;

PI Lal PG, Thornton M, Hafalia AJA, Yao MG, Nguyen DB, Gandhi AR;

PI Khan FA, Walhia NK, Griffin JA, Chinn AM, Elliott VS, Ramkumar J;

PI Arvizu CS, Forsythe J;

XX WPI; 2003-167112/16.

DR P-PSDB; ABP58227.

XX New human cell adhesion and extracellular matrix proteins, useful for
 diagnosing, treating or preventing autoimmune or inflammatory disorder
 (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,

cancer or hepatitis.

Claim 12; Page 168-170; 178pp; English.

The present sequence is that of Incyte clone 7326129CB1 cDNA encoding human cell adhesion and extracellular matrix protein 4 (CADECM-4). A representative cDNA library for the full-length polynucleotide is MUSLTR02, constructed from the muscle tissue RNA of a Caucasian adult man. Homology data suggest the encoded protein to a titin muscle protein. The invention provides CADECM-1 to -11 polypeptides (see ABP58224-34) and polynucleotides (see ABZ24578-88), expression vectors, host cells, antibodies, agonists and antagonists. These are useful for diagnosing, treating or preventing disorders associated with aberrant expression of CADECM, particularly cell proliferative disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, paroxysmal nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary thrombocytopaenia or cancer), developmental disorders (e.g. renal tubular acidosis, anaemia or mental retardation), neurological disorders (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), reproductive disorders (e.g. infertility or a disruption in the menstrual cycle), or autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral, bacterial, fungal, parasitic, protozoal or helminthic infections

Sequence 7564 BP; 1594 A; 1999 C; 2711 G; 1260 T; 0 U; 0 Other;

Query Match 99.7%; Score 997.8; DB 8; Length 7564;

Best Local Similarity 99.8%; Pred. No. 1.8e-230;

Matches 999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	GCCCTGTGGAGTGGAGAGGGGCCGAGAACCTCAGAGATGGGACAGATACATCTG	60
Db	4546	GCCCTGTGGAGTGGAGAGGGGCCGAGAACCTCAGAGATGGGACAGATACATCTG	4605
Qy	61	AGGACAGAGGGGACAGGTGTGAGCTGCAGATCTGTGGCTGGCCATGGGAGCGCGG	120
Db	4606	AGGACAGAGGGGACAGGTGTGAGCTGCAGATCTGTGGCTGGCCATGGGAGCGCGG	4665
Qy	121	GAGTACTTGTGTGTGCGGACAGAGAGGACCTCAGCAGCTCACCATCAGGGCTCTG	180
Db	4666	GAGTACTTGTGTGTGCGGACAGAGAGGACCTCAGCAGCTCACCATCAGGGCTCTG	4725
Qy	181	CTGCCAGGTTCTAGAAAGATGTGAAACACAGAGGCGCAGAGAGGGCCACCGCTGTG	240
Db	4726	CCTGCCAGGTTCTAGAAAGATGTGAAACACAGAGGCGCAGAGAGGGCCACCGCTGTG	4785
Qy	241	CTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGAGTGGAGAAAGGGGTCTGACACCTC	300
Db	4786	CTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGAGTGGAGAAAGGGGTCTGACACCTT	4845
Qy	301	AGAGATGGGACAGATACAGCTCTGAGGACGAGCGGACTAAATGTGAGCTGCAGATTCGT	360
Db	4846	AGAGATGGGACAGATACAGCTCTGAGGACGAGCGGACTAAATGTGAGCTGCAGATTCGT	4905
Qy	361	GGCTGGCCATGGCAGACACTGGGGAGTACTGCTGCGGTGCGGGCAGGAGACCTCG	420
Db	4906	GGCTGGCCATGGCAGACACTGGGGAGTACTGCTGCGGTGCGGGCAGGAGACCTCG	4965
Qy	421	GCTATGCTCACCGTCAAGGCTCTACCCATCAAGTTCAAGAGGGTCTGAGGAACGAAGAG	480
Db	4966	GCTATGCTCACCGTCAAGGCTCTACCCATCAAGTTCAAGAGGGTCTGAGGAACGAAGAG	5025
Qy	481	GCCACAGAAGGGGCAACAGCCGCTGCTGGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG	540
Db	5026	GCCACAGAAGGGGCAACAGCCGCTGCTGGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG	5085
Qy	541	TGGTGGAGGGGATGAGACCTCAGAGATGGAGACAGACACAGCCTTGAGCAGCAGCGG	600
Db	5086	TGGTGGAGGGGATGAGACCTCAGAGATGGAGACAGACACAGCCTTGAGCAGCAGCGG	5145

Qy	601	GCCAGGTGTGAGCTGCAGATCCGGCGCTCGTGGCAGAGGACGCTGGGGAGTACTCTGTC	660
Db	5146	GCCAGGTGTGAGCTGCAGATCCGGCGCTCGTGGCAGAGGACGCTGGGGAGTACTCTGTC	5205
Qy	661	ATGTGCGGGGAAGAGAGGAGGACCTCAGCCATGCTCACCGTCAGGGCCATGCCCTTCCAAGTTC	720
Db	5206	ATGTGCGGGGAAGAGAGGAGGACCTCAGCCATGCTCACCGTCAGGGCCATGCCCTTCCAAGTTC	5265
Qy	721	ATAGAGGCTGTGAGGATGAAGAGGCCACAGAGGGGACACGGCCACGCTGTGCTGTGAG	780
Db	5266	ATAGAGGCTGTGAGGATGAAGAGGGCCACAGAGGGGACACGGCCACGCTGTGCTGTGAG	5325
Qy	781	CTGAGCAAGGCGGCACCGGTGGAGTGGAGGAAGGGGCATGAGACCTTCAGAGATGGGGAC	840
Db	5326	CTGAGCAAGGCGGCACCGGTGGAGTGGAGGAAGGGGCATGAGACCTTCAGAGATGGGGAC	5385
Qy	841	AGACACAGCCTGAGGACGAGGACGGGTCCAGTGTGAGCTGCAGATCCGTGGCCTGGCTGTG	900
Db	5386	AGACACAGCCTGAGGACGAGGACGGGTCCAGTGTGAGCTGCAGATCCGTGGCCTGGCTGTG	5445
Qy	901	GTGATGCCGGGAGTACTCTGTGTGTGCGGCGAGGAGGACCTCAGCACAACCTCACT	960
Db	5446	GTGATGCCGGGAGTACTCTGTGTGTGCGGCGAGGAGGACCTCAGCACAACCTCACT	5505
Qy	961	GTGAGGGCCCTGCTGCTGCCAGATTCATAGAAGATGTGAAAAA	1001
Db	5506	GTGAGGGCCCTGCTGCTGCCAGATTCATAGAAGATGTGAAAAA	5546

RESULT 4
ABV99363
ID ABV99363 standard; DNA; 14061 BP.
XX AC ABV99363;
XX DT 27-JAN-2003 (first entry)
XX DE Human NOV13b coding sequence.
XX KW Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS;
KW antiinflammatory; cardiac; haemostatic; neuroprotective; anorectic;
KW neutropenic; immunosuppressive; osteopathic; antiparkinsonian; cancer;
KW antifertility; cerebroprotective; gene therapy; NOVX; NOV; fertility;
KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;
KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; cardiovascular disorder;
KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;
KW metabolic syndrome X; wasting disorder; cell differentiation; gene;
KW cell proliferation; haematopoiesis; wound healing; angiogenesis; ds.
XX OS Homo sapiens.
XX FN WO200272771-A2.
XX PD 19-SEP-2002.
XX PF 08-MAR-2002; 2002WO-US007288.
XX PR 08-MAR-2001; 2001US-0274101P.
PR 08-MAR-2001; 2001US-0274194P.
PR 08-MAR-2001; 2001US-0274281P.
PR 08-MAR-2001; 2001US-0274322P.
PR 09-MAR-2001; 2001US-0274849P.
PR 12-MAR-2001; 2001US-0275235P.
PR 13-MAR-2001; 2001US-0275578P.
PR 13-MAR-2001; 2001US-0275579P.
PR 14-MAR-2001; 2001US-0275601P.
PR 16-MAR-2001; 2001US-0276776P.
PR 19-MAR-2001; 2001US-0276994P.
PR 20-MAR-2001; 2001US-0277239P.
PR 20-MAR-2001; 2001US-0277321P.

PR 20-MAR-2001; 2001US-0277327P.
PR 20-MAR-2001; 2001US-0277338P.
PR 21-MAR-2001; 2001US-0277791P.
PR 22-MAR-2001; 2001US-0277833P.
PR 23-MAR-2001; 2001US-0278152P.
PR 26-MAR-2001; 2001US-0278894P.
PR 27-MAR-2001; 2001US-0278999P.
PR 27-MAR-2001; 2001US-0279038P.
PR 28-MAR-2001; 2001US-0279344P.
PR 30-MAR-2001; 2001US-0279995P.
PR 30-MAR-2001; 2001US-0280233P.
PR 02-APR-2001; 2001US-0280802P.
PR 02-APR-2001; 2001US-0280822P.
PR 02-APR-2001; 2001US-0280900P.
PR 04-APR-2001; 2001US-0281194P.
PR 13-APR-2001; 2001US-0283675P.
PR 30-APR-2001; 2001US-0287424P.
PR 03-MAY-2001; 2001US-0288066P.
PR 03-MAY-2001; 2001US-0288342P.
PR 03-MAY-2001; 2001US-0288528P.
PR 15-MAY-2001; 2001US-0291190P.
PR 16-MAY-2001; 2001US-0291099P.
PR 30-MAY-2001; 2001US-0291240P.
PR 30-MAY-2001; 2001US-0294485P.
PR 31-MAY-2001; 2001US-0294689P.
PR 31-MAY-2001; 2001US-0294899P.
PR 16-AUG-2001; 2001US-0312903P.
PR 10-SEP-2001; 2001US-0318462P.
PR 12-SEP-2001; 2001US-0318770P.
PR 27-SEP-2001; 2001US-0325430P.
PR 27-SEP-2001; 2001US-0325681P.
PR 18-OCT-2001; 2001US-0330380P.
PR 31-OCT-2001; 2001US-0335301P.
PR 14-NOV-2001; 2001US-0332172P.
PR 14-NOV-2001; 2001US-0332271P.
PR 14-NOV-2001; 2001US-0332272P.
PR 14-NOV-2001; 2001US-0333184P.
PR 14-NOV-2001; 2001US-0333272P.
PR 21-NOV-2001; 2001US-0332094P.
PR 03-DEC-2001; 2001US-0337428P.
PR 03-DEC-2001; 2001US-0338092P.
PR 04-DEC-2001; 2001US-0337185P.
PR 03-JAN-2002; 2002US-0345705P.
PR 08-MAR-2002; 2002US-00093463.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
PI Boldog FL, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Vernet CM;
PI Pena CEA, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;
PI Voss E, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
PI Taupier RJ, Padigar M, Shency SG, Kekuda R, Gusev VY, Pochart PF;
PI Zhong M;
XX
XX WPI; 2002-732824/79.
DR P-PSDB; ABP70085.
DR
XX
XX New NOVX polypeptides and polynucleotides, useful for preventing,
PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
PT Alzheimer's disease, dyslipidemia, obesity, immune or hematopoietic
PT disorders, and asthma.
XX
ES Claim 16; Page 138-142; 619pp; English.
XX
XX The present invention relates to new isolated proteins (NOVX) and their
CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
CC any number from 1 to 48. The NOVX proteins and coding sequences are
CC useful in the manufacture of a medicament for treating a syndrome

CC associated with a human disease, preferably a NOVX-associated disorder.
CC The NOVX coding sequences and proteins are useful for treating,
CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
CC obesity, infectious disease, anorexia, cancer-associated cachexia,
CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
CC disease, immune disorders, hematopoietic disorders, cardiovascular
CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
CC disturbances associated with obesity, metabolic syndrome X or wasting
CC disorders associated with chronic diseases or various cancers. The NOVX
CC coding sequences and proteins may also be used as targets for the
CC identification of small molecules that modulate or inhibit e.g.
CC neurogenesis, cell differentiation, cell proliferation, hematopoiesis,
CC wound healing and angiogenesis, in gene therapy, in generation of
CC antibodies that bind immunospecifically to NOVX substances for use in
CC therapeutic or diagnostic methods
XX
SQ Sequence 14061 BP; 2854 A; 4029 C; 4839 G; 2339 T; 0 U; 0 Other;
Query Match 99.7%; Score 997.8; DB 6; Length 14061;
Best Local Similarity 99.8%; Pred. No. 2.1e-230;
Matches 999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCCCCTGTGGAGTGGAGAGGGCCGAGAACCTTCAGAGATGGGGACAGATACATCTCG 60
DB 2661 GCCCCTGTGGAGTGGAGAGGGCCGAGAACCTTCAGAGATGGGGACAGATACATCTCG 2720
QY 61 AGGCAGAGGGGACCAAGTGTGAGCTGCAGATCTGTGGCTGGCCATGGCGACCCCGG 120
DB 2721 AGGCAGAGGGGACCAAGTGTGAGCTGCAGATCTGTGGCTGGCCATGGCGACCCCGG 2780
QY 121 GAGTACTTGTGTGTGGGGCAGGAGGAGGACCTCAGGCAGCTCACCATCAGGGCTCTG 180
DB 2781 GAGTACTTGTGTGTGGGGCAGGAGGAGGACCTCAGGCAGCTCACCATCAGGGCTCTG 2840
QY 181 CCTGCCAGTTTCATAGAAGATGTGAAAAACAGAGGCGCAGAGAGGGGCCACCGCTGTG 240
DB 2841 CCTGCCAGTTTCATAGAAGATGTGAAAAACAGAGGCGCAGAGAGGGGCCACCGCTGTG 2900
QY 241 CTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGAGTGGAGAGAAAGGGGTCTGAGACCTC 300
DB 2901 CTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGAGTGGAGAGAAAGGGGTCTGAGACCTC 2960
QY 301 AGAGATGGGGACAGATACAGCCTCAGGCAGGACGGGACTTAATGTGAGCTCAGATTCTG 360
DB 2961 AGAGATGGGGACAGATACAGCCTCAGGCAGGACGGGACTTAATGTGAGCTCAGATTCTG 3020
QY 361 GGCCTGGCCATGGCAGACACTTGGGGAGTACTCTGCGTGTGCGGGCAGGAGAGGACCTCG 420
DB 3021 GGCCTGGCCATGGCAGACACTTGGGGAGTACTCTGCGTGTGCGGGCAGGAGAGGACCTCG 3080
QY 421 GCTATGCTCACCGTCAGGGCTCTACCCATCAAGTTTCAGAGGGTCTTGAGGAACGAAGAG 480
DB 3081 GCTATGCTCACCGTCAGGGCTCTACCCATCAAGTTTCAGAGGGTCTTGAGGAACGAAGAG 3140
QY 481 GCCACAGAGGGGCAACAGCCGCTCTCGGCTGAGCTGAGCAAGAGTGGCCCCCGTGGAG 540
DB 3141 GCCACAGAGGGGCAACAGCCGCTCTCGGCTGAGCTGAGCAAGAGTGGCCCCCGTGGAG 3200
QY 541 TGGTGGAGGGGCATGAGACCTTCAGAGATGGAGACAGACACAGCCTCAGGCAGGACGGG 600
DB 3201 TGGTGGAGGGGCATGAGACCTTCAGAGATGGAGACAGACACAGCCTCAGGCAGGACGGG 3260
QY 601 GCCAGGTGTGAGCTGCAGATCCCGGGCTCTGTGCGAGAGGACGCTGGGGAGTACTGTGTC 660
DB 3261 GCCAGGTGTGAGCTGCAGATCCCGGGCTCTGTGCGAGAGGACGCTGGGGAGTACTGTGTC 3320
QY 661 ATGTGCGGGAGGAGGAGGACCTCAGCCATCTCTACCTCAGGGCCATGCTTCCAGTTTC 720
DB 3321 ATGTGCGGGAGGAGGAGGACCTCAGCCATGCTCTACCTCAGGGCCATGCTTCCAGTTTC 3380
QY 721 ATAGAGGTCTGAGGAATGAAGAGGCCACAGAGGGGACACGGCCACCTCTGTGGTGTGAG 780
DB 3381 ATAGAGGTCTGAGGAATGAAGAGGCCACAGAGGGGACACGGCCACCTCTGTGGTGTGAG 3440

CC wound healing and angiogenesis, in gene therapy, in generation of
CC antibodies that bind immunospecifically to NOVX substances for use in
CC therapeutic or diagnostic methods

XX SQ Sequence 14109 BP; 2862 A; 4045 C; 4854 G; 2348 T; 0 U; 0 Other;

Query Match 99.7%; Score 997.8; DB 6; Length 14109;
Best Local Similarity 99.8%; Pred. No. 2.1e-230;
Matches 999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCTGTGGAGTGGAGGAGGGGCCGAGAACCTCAGAGATGGGGACAGATACATCCTG 60
Db 2661 GCCCTGTGGAGTGGAGGAGGGGCCGAGAACCTCAGAGATGGGGACAGATACATCCTG 2720

Qy 61 AGCAGAGAGGGACACAGGTGTGAGCTGCAGATCTGTGGCTGGCCATGGCGGACGCCGG 120
Db 2721 AGCAGAGAGGGACACAGGTGTGAGCTGCAGATCTGTGGCTGGCCATGGCGGACGCCGG 2780

Qy 121 GAGTACTTGTGTGTGGCGGACAGAGAGGACCTCAGCCACGCTCACCATCAGGGCTCTG 180
Db 2781 GAGTACTTGTGTGTGGCGGACAGAGAGGACCTCAGCCACGCTCACCATCAGGGCTCTG 2840

Qy 181 CCTGCCAGGTTTCATAGAGATGTGAAAGAACAGAGAGGCCAGAGAGGGCCACGGCTGTG 240
Db 2841 CCTGCCAGGTTTCATAGAGATGTGAAAGAACAGAGAGGCCAGAGAGGGCCACGGCTGTG 2900

Qy 241 CTCAGTGTGAGCTGAACAGTGCAGCCCTGTGGAGTGGAGAAAGGGGTCTGAGACCTC 300
Db 2901 CTCAGTGTGAGCTGAACAGTGCAGCCCTGTGGAGTGGAGAAAGGGGTCTGAGACCTC 2960

Qy 301 AGAGATGGGACAGATACAGCTGAGGACGAGGACGGGACTTAATGTGAGCTCAGATTCTG 360
Db 2961 AGAGATGGGACAGATACAGCTGAGGACGAGGACGGGACTTAATGTGAGCTCAGATTCTG 3020

Qy 361 GGCCTGGCCATGCGACACACTGGGAGTACTCTGCTGTGGCGGACGAGAGGACCTCG 420
Db 3021 GGCCTGGCCATGCGACACACTGGGAGTACTCTGCTGTGGCGGACGAGAGGACCTCG 3080

Qy 421 GCTATGCTCACCGTCAGGCTCTACCCATCAAGTTTCACAGAGGGTCTCAGGAACGAAG 480
Db 3081 GCTATGCTCACCGTCAGGCTCTACCCATCAAGTTTCACAGAGGGTCTCAGGAACGAAG 3140

Qy 481 GCCACAGAGGGGCAACAGCCGTGCTGGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG 540
Db 3141 GCCACAGAGGGGCAACAGCCGTGCTGGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG 3200

Qy 541 TGGTGAAGGGCATGAGACCTCAGAGATGAGACAGACACAGCCTCAGGCAGGACGG 600
Db 3201 TGGTGAAGGGCATGAGACCTCAGAGATGAGACAGACACAGCCTCAGGCAGGACGG 3260

Qy 601 GCCAGGTGTGAGCTCAGATCCGCGCTCTCGTGGCAGAGAGCGCTGGGGAGTACCTGTGC 660
Db 3261 GCCAGGTGTGAGCTCAGATCCGCGCTCTCGTGGCAGAGAGCGCTGGGGAGTACCTGTGC 3320

Qy 661 ATGTCGGGAGGAGAGACCTCAGCATGCTCACCCTCAGGGCCATGCTTCCAAGTTC 720
Db 3321 ATGTCGGGAGGAGAGACCTCAGCATGCTCACCCTCAGGGCCATGCTTCCAAGTTC 3380

Qy 721 ATAGAGGTCTGAGCAATGAAGAGCCCAAGAGGGGACACGGCCACGCTGTGTGTGAG 780
Db 3381 ATAGAGGTCTGAGCAATGAAGAGCCCAAGAGGGGACACGGCCACGCTGTGTGTGAG 3440

Qy 781 CTGAGCAAGGGCGGACCGGTGTGAGTGGAGAGGGGCGATGAGACCTCAGAGATGGGAC 840
Db 3441 CTGAGCAAGGGCGGACCGGTGTGAGTGGAGAGGGGCGATGAGACCTCAGAGATGGGAC 3500

Qy 841 AGACACAGCTTGAAGCAGAACGGGTCCAGGTGTGAGCTGCAGATCCGTGGCCCTGGCTGTG 900
Db 3501 AGACACAGCTTGAAGCAGAACGGGTCCAGGTGTGAGCTGCAGATCCGTGGCCCTGGCTGTG 3560

Qy 901 GTGGATCCGGGGAGTACTCTGTGTGTGTCGGGACAGGAGACCTCAGCCACACTCACT 960
Db 3561 GTGGATCCGGGGAGTACTCTGTGTGTGTCGGGACAGGAGACCTCAGCCACACTCACT 3620

Qy 961 GTGAGGCGCTGCTGCCAGATTCATACAGATCTGAAAA 1001
Db 3621 GTGAGGCGCTGCTGCCAGATTCATACAGATCTGAAAA 3661

RESULT 6

AAF64470/C

ID AAF64470 standard; cDNA; 352 BP.

XX AC AAF64470;

XX DT 09-APR-2001 (first entry)

XX DE Novel human polynucleotide, SEQ ID NO: 226.

XX KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
breast cancer; lung cancer; cancer detection; ss.

XX OS Homo sapiens.

XX FN WO200102568-A2.

XX PD 11-JAN-2001.

XX PF 30-JUN-2000; 2000WO-US018374.

XX PR 02-JUL-1999; 99US-0142310P.

XX PR 02-JUL-1999; 99US-0142311P.

XX PA (CHIR) CHIRON CORP.
(HYSE-) HYSEQ INC.

XX PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;
Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;
Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;
Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;

XX DR WPI; 2001-091805/10.

XX PT Library of polynucleotides for diagnosing a cancerous state of a
mammalian cell and detecting cancer, particularly of the colon or
prostate, comprises 3351 human polynucleotide sequences.

XX PS Claim 9; Page 576; 1046pp; English.

XX CC The present sequence is one of 3351 sequences in a library of human
polynucleotides. The library is used to detect differentially expressed
genes correlated with a cancerous state of a mammalian cell and can
detect colon, prostate, breast and lung cancer. The library can be used
to produce probes for detection of mRNA and to produce additional copies
of the polynucleotides. The probes can be used for chromosome mapping of
the polynucleotide and for detection of transcription levels. Ribozymes
or antisense oligonucleotides can be generated. The polynucleotides and
their gene products are used as genetic or biochemical markers (e.g. in
blood or tissues) that will detect the earliest changes along the
carcinogenesis pathway and/or monitor the efficacy of therapies and
preventive interventions. The polynucleotides, polypeptides and
antibodies against them can be used in pharmaceutical compositions to
treat the cancers and proliferative disorders such as neoplasia,
dysplasia and hyperplasia

SQ Sequence 352 BP; 56 A; 138 C; 84 G; 73 T; 0 U; 1 Other;

Query Match 34.3%; Score 343.4; DB 5; Length 352;

Best Local Similarity 99.7%; Pred. No. 5.1e-73;

Matches 344; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 631 GTGGCAGAGGACGGTGGGGAGTACCTGTGCTATGTGCGGGAAGGAGGACCTCAGCCATG 690

Db 351 GTGGCAGAGGACGGTGGGGAGTACCTGTGCTATGTGCGGGAAGGAGGACCTCAGCCATG 292

Qy 691 CTCACCGTCAGGGGCCATGCCCTTCAGTTCTAGAGGCTCTAGAGGAGGCCACA 750

|||||
291 CTCACCGTCAAGGCGCATGCTTCAAGTTTCATAGAGGGTCTGAGGAATGAAGAGGCCACA 232
|||||
751 GAAGGGGACACGGCCACGCTGTGTGTGAGCTGAGCAAGGCGCACCGGTGGAGTGGAGG 810
|||||
231 GAAGGGGACACGGCCACGCTGTGTGTGAGCTGAGCAAGGCGCACCGGTGGAGTGGAGG 172
|||||
811 AAGGGGCATGAGACCCCTCAGAGATGGGGACACACAGCCTGAGGCAGGACGGGTCCAGG 870
|||||
171 AAGGGGCATGAGACCCCTCAGAGATGGGGACACACAGCCTGAGGCAGGATGGGTCCAGG 112
|||||
871 TGTGAGCTGCAGATCCGTGGCTGTGTGTGATGCCGGGGAGTACTGTGTGTGTGC 930
|||||
111 TGTGAGCTGCAGATCCGTGGCTGTGTGTGATGCCGGGGAGTACTGTGTGTGTGC 52
|||||
931 GGGCAGGAGAGGACCTCAGCCACACTCACTCTCAGGGCCCTGCCT 975
|||||
51 GGGCAGGAGAGGACCTCAGCCACACTCACTGTGAGGGCCCTGCCT 7

RESULT 7
ABQ61169
ID ABQ61169 standard; cDNA; 642 BP.
XX
AC ABQ61169;
XX
DT 26-FEB-2003 (first entry)
XX
DE Obscurin OBSCN gene encoding sequence.

Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;
gene therapy; nutritional supplement; wound; burn; ulcer;
Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
autoimmune disorder; inflammation; vulnery; gene; ss.

OS Homo sapiens.

PN WO200231111-A2.

PD 18-APR-2002.

PF 11-OCT-2001; 2001WO-US027760.

PR 12-OCT-2000; 2000US-00687527.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

DR WPI; 2002-426278/45.

DR N-PSDB; ABP43925.

XX New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.

PS Claim 1; SEQ ID # 382; 357pp + Sequence Listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences. The
CC activity of polynucleotides of the invention may be described as,
CC vulnery, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are
CC useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records ABQ60788-
CC ABQ61233 represent polynucleotides of the invention. NOTE: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 642 BP; 147 A; 169 C; 205 G; 121 T; 0 U; 0 Other;

Query Match 26.5%; Score 265; DB 6; Length 642;
Best Local Similarity 100.0%; Pred. No. 4.6e-54;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 439 GCTTACCCATCAAGTTTCACAGAGGGTCTGAGGAACGAAGAGGCCACAGAAGGGGCAACA 498
|||||
Db 227 GCTTACCCATCAAGTTTCACAGAGGGTCTGAGGAACGAAGAGGCCACAGAAGGGGCAACA 286
|||||
Qy 499 GCCGTGTCGGGTGTGAGTGTGAGCAAGATGCCCCCTGTGAGTGTGTGAAGGGGCATGAG 558
|||||
Db 287 GCCGTGTCGGGTGTGAGTGTGAGCAAGATGCCCCCTGTGAGTGTGTGAAGGGGCATGAG 346
|||||
Qy 559 ACCCTCAGAGATGGAGACACACAGCCTGAGGCAGGACGGGCCAGGTGTGAGCTGCAG 618
|||||
Db 347 ACCCTCAGAGATGGAGACACACAGCCTGAGGCAGGACGGGCCAGGTGTGAGCTGCAG 406
|||||
Qy 619 ATCCGGCGCCTCGTGGCAGAGGACCGCTGGGGAGTACCTGTGCATGTGCGGGAAGGAGAGG 678
|||||
Db 407 ATCCGGCGCCTCGTGGCAGAGGACCGCTGGGGAGTACCTGTGCATGTGCGGGAAGGAGAGG 466
|||||
Qy 679 ACCTCAGCCATGCTCACCGTCAGGG 703
|||||
Db 467 ACCTCAGCCATGCTCACCGTCAGGG 491

RESULT 8

ABQ61180

ID ABQ61180 standard; cDNA; 707 BP.

XX AC ABQ61180;

DT 26-FEB-2003 (first entry)

XX Obscurin OBSCN gene #2 encoding sequence.

XX Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;
XX gene therapy; nutritional supplement; wound; burn; ulcer;
XX Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
XX autoimmune disorder; inflammation; vulnery; gene; ss.

OS Homo sapiens.

PN WO200231111-A2.

PD 18-APR-2002.

PF 11-OCT-2001; 2001WO-US027760.

PR 12-OCT-2000; 2000US-00687527.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

DR WPI; 2002-426278/45.

DR N-PSDB; ABP43936.

XX New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.

PS Claim 1; SEQ ID # 393; 357pp + Sequence Listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences. The
CC activity of polynucleotides of the invention may be described as,
CC vulnery, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are

CC useful for treating a mammalian subject, or as nutritional sources or
 CC supplements. These are useful in gene therapy, particularly for treating
 CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
 CC inflammation. The nucleic acids and polypeptides are also useful in
 CC diagnostic and research methods. The sequences given in records AB060788-
 CC AB061233 represent polynucleotides of the invention. NOTE: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 707 BP; 154 A; 198 C; 213 G; 142 T; 0 U; 0 Other;
 Query Match 26.5%; Score 265; DB 6; Length 707;
 Best Local Similarity 100.0%; Pred. No. 4.7e-54;
 Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 439 GCTCTACCCATCAAGTTCCACAGAGGGTCTGAGGAACGAGAGGCCACAGAGGGGCCAACA 498
 DB 227 GCTCTACCCATCAAGTTCCACAGAGGGTCTGAGGAACGAGAGGCCACAGAGGGGCCAACA 286
 QY 499 GCCGTGTGCGGTGTGAGCTGACGACGACGATGCCCCCGTGGAGTGGTGGAGGGGCATGAG 558
 DB 287 GCCGTGTGCGGTGTGAGCTGACGACGATGCCCCCGTGGAGTGGTGGAGGGGCATGAG 346
 QY 559 ACCCTCAGATGAGACAGACACAGCTTGAGGACGAGCGGGCCAGGTGTGAGCTGCAG 618
 DB 347 ACCCTCAGATGAGACAGACACAGCTTGAGGACGAGCGGGCCAGGTGTGAGCTGCAG 406
 QY 619 ATCCGCGCCCTCGTGGCAGAGCGCTGGGAGTACCTGTGCATGTGCGGGAAGGAGG 678
 DB 407 ATCCGCGCCCTCGTGGCAGAGCGCTGGGAGTACCTGTGCATGTGCGGGAAGGAGG 466
 QY 679 ACCTCAGCCATGCTCACCGTCAGG 703
 DB 467 ACCTCAGCCATGCTCACCGTCAGG 491
 RESULT 9
 ABK99965
 ID ABK99965 standard; DNA; 2155 BP.
 AC ABK99965;
 XX
 XX 21-OCT-2002 (first entry)
 XX
 DE DNA encoding human secreted protein SCEP-48.
 XX
 KW Secreted protein; SCEP; human; cell proliferative disorder; cancer;
 KW keratosis; arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;
 KW psoriasis; autoimmune disorder; inflammatory disorder; AIDS; arthritis;
 KW acquired immunodeficiency syndrome; adult respiratory distress syndrome;
 KW Addison's disease; allergy; asthma; osteoporosis; autoimmune thyroiditis;
 KW Crohn's disease; dermatitis; diabetes; Graves' disease; haemodialysis;
 KW glomerulonephritis; scleroderma; systemic lupus erythematosus; uveitis;
 KW systemic sclerosis; ulcerative colitis; infection; trauma; Pick disease;
 KW cardiovascular disorder; congestive heart failure; angina; epilepsy;
 KW hypertensive heart disease; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; amyotrophic lateral sclerosis; stroke; dementia;
 KW metabolic disorder; endocrine disorder; toxic myopathy; mental disorder;
 KW schizophrenic disorder; glaucoma; sensorineural hearing loss; cataract;
 KW transgenic animal; gene; da.
 XX
 OS Homo sapiens.
 XX
 PN W0200248337-A2.
 XX
 XX 20-JUN-2002.
 PD
 XX 12-DEC-2001; 2001WO-US048517.
 PF
 XX

PR 13-DEC-2000; 2000US-0255639P.
 PR 21-DEC-2000; 2000US-0257852P.
 PR 05-JAN-2001; 2001US-0260105P.
 PR 18-JAN-2001; 2001US-0262932P.
 PR 18-JAN-2001; 2001US-0263096P.
 PR 19-JAN-2001; 2001US-0263090P.
 PR 02-FEB-2001; 2001US-0265926P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 PI Griffin JA, Yao MG, Duggan BM, Yue H, Ding L, Lal PG, Lee EA;
 PI Ramkumar J, Thangavelu K, Xu Y, Lee S, Tang YT, Nguyen DB;
 PI Warren BA, Honcheill CD, Gietzen KJ, Baughn MR, Gandhi AR, Arvizu C;
 PI Walia NK, Lu Y, Elliott VW, Lu DAM, Hafalia AUA, Azimzai Y;
 PI Khan FA, Tran UK;
 XX WPI; 2002-583509/62.
 DR P-PSDB; ABG69668.
 XX
 PT Novel human secreted proteins and polynucleotides for diagnosing,
 PT preventing or treating cell proliferative, autoimmune/inflammatory,
 PT cardiovascular, neurological and developmental disorders.
 XX
 PS Claim 5; Page 230; 234pp; English.
 XX
 CC The invention describes an isolated polypeptide chosen from secreted
 CC proteins (I), SCEP 1-54. (I) and the polynucleotide encoding it (II) are
 CC useful for screening a compound for effectiveness as an agonist or
 CC antagonist of (I) or compound that alters expression of (II). (I), the
 CC identified agonist and antagonist are useful for treating a disease or
 CC condition associated altered expression of functional SCEP in a patient.
 CC An antibody specific to (I) is useful for detecting the presence of (I).
 CC Purifying (I) from a sample and for diagnosing a condition or disease
 CC associated with expression of SCEP in a subject or in a biological
 CC sample. (I) and (II) and modulators of (I) are useful for diagnosis,
 CC treatment and prevention of cell proliferative disorders (e.g. cancer,
 CC keratosis, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis and
 CC psoriasis), autoimmune/inflammatory disorders (e.g. acquired
 CC immunodeficiency syndrome (AIDS), adult respiratory distress syndrome,
 CC Addison's disease, allergies, asthma, osteoporosis, autoimmune
 CC thyroiditis, Crohn's disease, dermatitis, diabetes, Graves' disease,
 CC glomerulonephritis, arthritis, scleroderma, systemic lupus erythematosus,
 CC bacterial, fungal, parasitic, protozoal, helminthic infections and
 CC systemic sclerosis, ulcerative colitis, haemodialysis, uveitis; viral,
 CC trauma), cardiovascular disorders (e.g. congestive heart failure, angina,
 CC hypertensive heart disease), neurological disorders (e.g. Alzheimer's and
 CC Pick disease, Parkinson disease, amyotrophic lateral sclerosis, epilepsy,
 CC stroke, Huntington's disease, multiple sclerosis, dementia, neuromuscular
 CC disorders, metabolic, endocrine and toxic myopathies, mental disorders,
 CC schizophrenic disorders, and developmental disorders (e.g. anaemia,
 CC epilepsy, hypothyroidism, glaucoma, sensorineural hearing loss and
 CC cataract). (II) is useful for creating transgenic animals to model human
 CC disease and to detect and quantify gene expression in biopsied tissues in
 CC which expression of SCEP is correlated with disease. This sequence
 CC encodes a human secreted protein (SCEP)
 XX
 SQ Sequence 2155 BP; 488 A; 530 C; 545 G; 592 T; 0 U; 0 Other;
 Query Match 23.9%; Score 238.8; DB 6; Length 2155;
 Best Local Similarity 59.8%; Pred. No. 1.3e-47;
 Matches 531; Conservative 0; Mismatches 282; Indels 95; Gaps 4;
 QY 175 GCTCTGCTGCCAGGTTTCATAGAGATGTGAAAACCCAGGAGGCCAGAGAGGGGCCACG 234
 DB 205 GCCCTGCTGCCAGGTTTCACAGAGGGTCTGAGGAATGAAGAGGCCATGAAGGGGCCACA 264
 QY 235 GCTGTGCTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGGAGTGGAGAAAGGGTCTGAG 294
 DB 265 GCCACACTGCAATGTGAGCTGAGCAAGGCAGCCCTCTGTGGAGTGGAGAAAGGGCTTGAG 324
 QY 295 ACCCTCAGATGGGACAGATACAGCTGAGGACGAGCGGACTTAAATGTGAGCTGCAG 354
 DB 325 GCTCTCAGATGGGGACAATAATACAGCTTGAGACAAGCGGGCTGTGTGTGAGCTGCAG 384

XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuho Y;
XX WPI: 2003-723558/69.
DR P-PSDB; ADM05704.
XX New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX Claim 1; SEQ ID NO 1946; 305pp; English.
XX The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC cDNA sequence of the invention.
XX SQ Sequence 2534 BP; 551-A; 679 C; 778 G; 526 T; 0 U; 0 Other;
Query Match 10.7%; Score 107.4; DB 11; Length 2534;
Best Local Similarity 47.7%; Pred. No. 6.2e-16;
Matches 420; Conservative 0; Mismatches 446; Indels 15; Gaps 3;
Qy 136 TCGCGGCGAGGAGACCTCAGCAGCTCACCATCAGGGCTCTGCTCGCCAGGTTTCATA 195
Db 199 TCCCTGAAGTGTGCGACTCTGCCCGCTGAGAGTCCGAGTGAAGCCGGTGTGTCCTG 258
Qy 196 GAAGATGTGAAAACACGAGGCGCCAGAGAAAGGGCCAGCGCTGTCTCAGTGTGAGCT- 254
Db 259 AAGCGCTGGATGACCTGTCCGAGAGGAGCGCGCACCTTGCCCTCAGTGTGAAGTC 318
Qy 255 --GAACAGTCAGGCCCCCTGTGAGTGGAGAAAGGGGTCTGAGACCTTCAGAGATGGGGAC 312
Db 319 TCTGACCCGAGGCCCATCTGTGTGGCGCAAGATGCGGTGCGCTGGGCCCCAGTGAC 378
Qy 313 AGATACAGCTGAGGAGGAGCGGACTAAATGTGAGCTGAGATCTGTTGGCTGGCCCATG 372
Db 379 AAGTATGACTTCTGACACGCGCGGCGCACGCGGGGGCTCGTGGTGCATGACGTGAGCCCT 438
Qy 373 GCAGACACTGGGGAGTACTGTGTGTGTGCGGCGAGAGAGGACCTCGGCTATGCTACC 432
Db 439 GAAGACGCGCGCTGTACACTTGCACGCTGGGCTCCGAGGAGACCCGGCGCGGGTCCGC 498
Qy 433 GTCAGGGCTTACCCATCAAGTTTACAGAGGGTCTGAGGAACGAAGAGGCCACAGAAGGG 492
Db 499 GTGCACGATCTGCAGTGGGCATCACCAGAGGCTGAAGACATGAGAGTGTCTGAAGGG 558
Qy 493 GCAACAGCGCTGTCGGTGTGAGTGAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 552
Db 559 GAAAGCTGCAGCTTGTGAGTGGGTCTGTGTCACAGAGAGTGCAGGAGCCCGCCCATGTGG 618
Qy 553 CATGAGACCTTCAGAGATGGAGACACACAGCCCTGAGGAGGAG- - - - -GGGGCC 603
Db 619 ACAGTGGTGGAGAGACAGTGGGAGCTTCCAGGCGCTTCCAGGCCACACGTGAGGGCCGA 678
Qy 604 AGGTGTGAGCTGCAGATCCGGGGCTCTGTGCGAGAGGAGCGTGGGGAGTACTGTGCATG 663
Db 679 AAATACATCTCTGTTGTCGGGAGGCTGCACCAAGTGTATGCGGGAGTGTCTTCTCT 738
Qy 664 TCGCGGAGGAGAGGAGCCTCAGACCATGCTACCGTCAAGGCGCATGCTTCCAGATTCATA 723
Db 739 GTGCGGGGCGCTCAGCTTCAAGGCTCTCACTATTGTTCAGAGAGAGGCGCGCCACCATCATC 798
Qy 724 GAGGCTGTGAGGATGAGAGGCCACAGAGGGGACACGCGCTGTGTGTGAGCTG 783

Db 799 AAGCCCTGGNAGACCAGTGGTGGCGCCAGGGAGGAGCTGGAGCTGGCTGTGAGCTG 858
Qy 784 AGCAAGCGG--CACCGGTGAGTGGAGGAAGGGGCATGAGACCTTCAGAGATGGGAC 840
Db 859 TCACGGCGGGGAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 918
Qy 841 AGACACAGCCTGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db 919 AAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978
Qy 901 GTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 979 AAGGACGCGGGGAGTACACGCTGTGAGGTGAGGCTTCCAGAGACACAGCCGCTCCAT 1038
Qy 961 GTACAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1001
Db 1039 GTGGAAGAAAAGCAAACTGCTTCCAGAGAGGAGCTGACCA 1079
RESULT 12
ID ABN21414 standard; cDNA; 1005 BP.
AC ABN21414;
XX 24-JUN-2002 (first entry)
DT Human ORFX polynucleotide sequence SEQ ID NO:11305.
DE Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis; gene; ss.
OS Homo sapiens.
XX WO200192523-A2.
PN 06-DEC-2001.
PD 29-MAY-2001; 2001WO-US010836.
XX 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX (CURA-) CURAGEN CORP.
PA Shimkets RA, Leach MD;
PI WPI: 2002-106308/14.
XX P-PSDB; ABP05662.
DR Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders.
PS Disclosure; SEQ ID NO 11305; 1037pp; English.
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ

```
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1005 BP; 209 A; 303 C; 340 G; 152 T; 0 U; 1 Other;

Query Match      8.6%; Score 86; DB 6; Length 1005;
Best Local Similarity 47.4%; Pred. No. 7.2e-11;
Matches 325; Conservative 0; Mismatches 355; Indels 6; Gaps 2;

Qy 112 GACGCGGGAGTACTTGTGTGTGTCGGGCGAGGAGGACCTCAGCCACGCTCACCATC 171
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 172 AGGCTCTGCTGCCAGGTTCATAGAAGATGTGAAAAACAGAGGCGCAGAGAGGGGCC 231
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 325 AGAGAGAGCGCGCGCCCATCATCAAGCCCTCGAAGACCAAGTGGGTGGCGCCAGGGAG 384
Qy 232 ACGCTGTGCTGAGTGTGAGCTGAACAGTGC---AGCCCTGTGGAGTGAGAGAAAGGG 288
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 385 GACGTGAGCTGCGCTGTGAGCTGTCAAGCGCGGGAAGCGCCGTGCACCTGCTGAAGAC 444
Qy 289 TCTGAGACCTTCAGAGATGGGACAGATACAGCTGTAGGCGAGGACGGAATGTGTAG 348
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 445 GGAAGGCCATCCGACAGAGCCAGATATGATGTGTCTCGAGGGCAGATGGCCATG 504
Qy 349 CTGAGATTGCTGCGCTGGCCATGGCAGACTGGGAGTACTCGTGGTGTGCGGGCAG 408
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 505 CTGCTCATCCGCGGGGCTCGCTCAAGGACGCGGGCGAGTACACGTGTGAGGTGGAGCT 564
Qy 409 GAGAGGACCTCGGTATGCTCACCGTCAAGGCTTACCCATCAAGTTTCACAGAGGTTCTG 468
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 565 TCCAAGAGCAGCGCAGCTCCATGTGGAAAGAAAGCACTCTTCACAGAGGAGCTG 624
Qy 469 AGGAACGAAGAGGCCACAGAGGGGCAACAGCCGTGTGCGGTGTGA---GCTCAGCAAG 525
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 625 ACCAATCTCAGGTGGAGGAGAAAGGCACAGCTGTGTTACGTGCAAGACGGACACCC 684
Qy 526 ATGCCCCCGTGGAGTGTGAAGGGGATGAGACCTTCAGAGATGGAGACAGACAGCAGC 585
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 685 GCGGCCACAGTGACCTGGCGCAAGGGCTCTTGGAGCTACGGGCTCAGGGAAGCACAG 744
Qy 586 CTGAGGACGAGCGGGGCCAGTGTGAGTGCAGATCCGGGCTCTGTTGGCAGAGGAGCT 645
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 745 CCCAGCCAGGAGGGCCCTGACCTTCGCGCTCACCATGATGCTTGGAGAGGCGACAGC 804
Qy 646 GGGAGTACCTGTGATGTGGGGAAGGAGAGGACCTCAGCCATGCTCACCGTCAGGGCC 705
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 805 GACACCTATACCTGCGACATTGGCCAGGCCAGTCCGGGCCAGCTCTTAGTCCAGGC 864
Qy 706 ATGCTTTCCAGTTTCATAGAGGGTCTGAGGAATGAAGAGGCCACAGAGGGGACACGGCC 765
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 865 CGGAGAGTGACATCATCAGGAGCTGGAGGATGTGGATGTACAGGAGGGCTCTCGGCC 924
Qy 766 ACGTGTGGTGTGAGCTGAGCAAGGC 791
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 925 ACCTTCGCTTGGCGGATCTCCCGGCC 950

RESULT 13
ADM01849
ID ADM01849 standard; cDNA; 3935 BP.
XX
```

```
AC ADM01849;
XX
XX 20-MAY-2004 (first entry)
XX
XX Human cDNA of the invention SEQ ID NO:534.
XX
XX ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
XX Homo sapiens.
XX
XX EPI347046-A1.
XX
XX 24-SEP-2003.
XX
XX 12-APR-2002; 2002EP-00008400.
XX
XX 22-MAR-2002; 2002JP-00137785.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX WPI; 2003-723558/69.
XX P-PSDB; ADM04292.
XX
XX New polynucleotides and polypeptides are useful in gene therapy, for
XX developing a diagnostic marker or medicines for regulating their
XX expression and activity, or as a target of gene therapy.
XX
XX Claim 1; SEQ ID NO 534; 305pp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded
XX polypeptide. A polynucleotide of the invention may have a use in gene
XX therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
XX as a primer for synthesizing the polynucleotide or as a probe for
XX detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
XX useful in gene therapy, for developing a diagnostic marker or medicines
XX for regulating their expression and activity, or as a target of gene
XX therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
XX are useful as pharmaceutical agents. The present sequence represents a
XX cDNA sequence of the invention.
XX
SQ Sequence 3935 BP; 1032 A; 1041 C; 1145 G; 717 T; 0 U; 0 Other;

Query Match      6.2%; Score 62; DB 11; Length 3935;
Best Local Similarity 46.0%; Pred. No. 6.1e-05;
Matches 295; Conservative 0; Mismatches 335; Indels 12; Gaps 2;

Qy 309 GGACAGATACAGCTGTAGGCGAGGACGGGACTAAATGTGAGCTGCAGATTCTGGCCCTGGC 368
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1033 GGGCAAGTACGATGTGAAGCAGATGGGCACCAAGTACATGCTGTTATTAGCAAGTGAA 1092
Qy 369 CATGGCAGACACTGGGAGTACTCGTGGTGTGCGGCAGAGAGACCTCGGTATGCT 428
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1093 CATGAACGATGCTGGCATCTACAGCCCTGTCGTTGGGCGATAAGCGGATGAGTCAGAGCT 1152
Qy 429 CACCGTCAGGCTCTACCCATCAAGTTTCACAGAGGTTCTGAGNAACGAAGAGGCCACAGA 488
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1153 CACAGTCTGATGAGCCACTGAAGTTTCCTGGAGAGATGAAGCTCTGTGAGGTGACAGA 1212
Qy 489 AGGGGCAACAGCGCTGCTGCGGTGTGAGCTGTGAGCAAGATGGCCCCCTGGAGTGGTGGAA 548
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1213 GCGCCAGACAGCTGTGTTTGAGATCCGCTCTCCAGAAAGAGCCCACTTTGTGTGAA 1272
Qy 549 GGGGCATGAGACCTCTCAGAGATGGAGACAGACA-----CAGCCTGAGGAGGACGG 599
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1273 GTTCAATGGGAAGAGCTGAAGAGGGATGACAAGTATGAATCACGGTGTCCGAAGATGG 1332
Qy 600 GGCCAGGTGTGAGTGCAGATCCGGCCTCGTGGCAGAGAGCGCTGGGAGTACCTGTG 659
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1333 TCTGACGCAACACGCTTAAGATTAAAGATGCAGACTCAGTGCAGCGCGGAGTTCTCTGC 1392
```

Qy 660 CATGTGCGGGAAGAGAGACCTCAGCCATCTCACCCTCAGGCGCATGCTTCAAGTT 719
Dy 1393 TGAGCGGGGAACCTGTGTACAAAGGCGCCAGCTCACTATTACCGCATCCCCATCAAGTT 1452
Qy 720 CATAGAGGCTCTGAGGAATGAAGAGGCCACAGAGGGGACACGGCCACGCTGTGTGTGA 779
Dy 1453 TGTGAGCACTCAAAATGTACGTGTGAAGAGAGGAGTCCGCGATGCTGTGAGTGTGA 1512
Qy 780 GCTGA---GCAAGCGGCGACCGGTGGAGTGAAGAGGGGCGATGAGACCTTCAGAGATGG 836
Dy 1513 GCTGACATCCCAAGATGTGACACTGCGCTGGAAGAGGATGGCGAGCTGTGATGCTATGG 1572
Qy 837 GGACAGACACAGCTGAGCGAGGACGGGTCCAGGTGTGAGTGTGAGATCCCGTGGCTGGC 896
Dy 1573 CACTAAGTACAGCATGAACCATGAGGGGCAAGCAAGCAGAGCTGATCATCGAGGATGCACA 1632
Qy 897 TGTGTGATGCCGGGAGTACTCGTGTGTGTGCGGGCAGGA 938
Dy 1633 GCTCAGTGTGTTGGCGAGTACACTGTGTGTGGCCATGCAGGA 1674

RESULT 14

ABN40768
ID ABN40768 standard; DNA; 60 BP.

XX AC ABN40768;

XX DT 15-JUL-2002 (first entry)

XX DE Human spliced transcript detection oligonucleotide SEQ ID NO:13516.

XX KW Human; mouse; rat; splice transcript; detection; RNA transcript;

XX KM splice variant; transcriptome; oligonucleotide library; ss.

XX OS Homo sapiens.

XX PN WO200210449-A2.

XX PD 07-FEB-2002.

XX PF 20-JUL-2001; 2001WO-IB001903.

XX PR 28-JUL-2000; 2000US-0221607P.

XX PR 02-MAY-2001; 2001US-0287724P.

XX PA (COMP-) COMPUTEN INC.

XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX PS WPI; 2002-257383/30.

XX PT New oligonucleotide libraries comprising oligonucleotides which selectively hybridize to mRNAs transcribed from a transcription unit of a genome, useful for detecting tissue-, pathology-, and developmental-specific genes.

XX PS Example 1; SEQ ID NO 13516; 47pp; English.

XX CC The present invention describes oligonucleotide libraries for detecting messenger RNAs that populate a (sub-)transcriptome, where the (sub-)transcriptome comprises messenger RNAs transcribed from multiple transcription units that populate a genome. The library comprises several oligonucleotides, each capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, which encodes one or more messenger RNA splice variants. The oligonucleotide libraries are useful for detecting mRNAs from a biological sample, in expression profiling studies, in qualitatively or quantitatively characterizing the corresponding transcriptome, and in detecting RNA transcripts and splice variants of human or animal transcriptomes. The libraries may also be used as specialised mini libraries to detect transcripts of a sub-transcriptome under a particular biological or pathological state, and so allowing the detection of tissue

CC - and pathology-specific genes such as those genes only expressed in specific tissue under a specific pathological condition; to detect developmental specific genes; and to detect RNA transcripts and splice variants of a transcriptome of a patient suffering from a particular disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from rats, humans and mice, which are used in the exemplification of the present invention. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 60 BP; 18 A; 13 C; 20 G; 9 T; 0 U; 0 Other;

Query Match 6.0%; Score 60; DB 6; Length 60;

Best Local Similarity 100.0%; Pred. No. 7e-05;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 703 GCCATGCTTCCCAAGTTTCATAGAGGGTCTGAGGAATGAAGAGGCCACAGAGGGGACACG 762
Dy 1 GCCATGCTTCCCAAGTTTCATAGAGGGTCTGAGGAATGAAGAGGCCACAGAGGGGACACG 60

RESULT 15

ABX13540
ID ABX13540 standard; DNA; 93801 BP.

XX AC ABX13540;

XX DT 05-JUN-2003 (first entry)

XX DE Human RGS11 DNA.

XX KW RGS11; human; screening; cardiant; antianginal; gene therapy; gene; heart disorder; cardiac ischaemia; heart failure; angina; db.

XX OS Homo sapiens.

XX PN WO2002103355-A1.

XX PD 27-DEC-2002.

XX PF 17-JUN-2002; 2002WO-JP006019.

XX PR 18-JUN-2001; 2001JP-00183038.

XX PA (TAKE) TAKEDA CHEM IND LTD.

XX PI Koyama N, Tanida S, Yamamoto K;

XX DR WPI; 2003-167557/16.

XX PS P-PSDB; ABG74786.

XX PT Screening compounds regulating RGS11 expression and activity for prevention and treatment of heart disease.

XX PS Example 1; Page 262-316; 321pp; Japanese.

XX CC This invention describes a novel method for screening compounds for their ability to regulate the activity and expression of human RGS11 and its partial peptides and salts, by observing the expression or activity of RGS11 in the presence or absence of the test compound. The products of the invention have cardiant and antianginal activity and can be used for gene therapy. The methods and compositions are useful in the prevention, treatment and diagnosis of heart disorders such as cardiac ischaemia, heart failure and angina. This sequence encodes the human RGS11 protein described in the disclosure of the invention

XX SQ Sequence 93801 BP; 30092 A; 19478 C; 21592 G; 22639 T; 0 U; 0 Other;

Query Match 5.9%; Score 59; DB 9; Length 93801;

Best Local Similarity 43.4%; Pred. No. 0.00067;

Matches 376; Conservative 0; Mismatches 485; Indels 6; Gaps 2;

Qy 141 GCAGGAGGAGACCTCAGCCAGGCTCACCATCAGAGGGTCTGCCTCCAGGTTTCATAGAAGA 200

Db	31821	GGAGCTGAGCTCCACAGCACACAGCTGAAGGTCTTTAGAGGCCGATCCCTACTTCACTGTGAA	31880
Qy	201	TGTGAAAAACCAGGAGCCAGAGAGGGGCCACGGCTGTGCTGCAGTGTGAGCTGAACAG	260
Db	31881	ATTACATGACAAAATCGCTAGTGGAGAGGATGAGATTCTTTGAGTGTGAAGTGAGCA	31940
Qy	261	TGCAGCCCTGTGGAGTGGAGAAAGGGTCTGAGACCTCAGAGATGGGACAGATACAG	320
Db	31941	AGATGTACCAGTGAATGGTTCAAAGATGGTGAAGAGATTGTCCCTTCAACCCAAATATC	32000
Qy	321	CCTGAGCAGCAGCGGACTAAATGTGAGCTGCAGATTCTGGCTTGGCCATGGCAGACAC	380
Db	32001	TATCAAGGCAGATGGCCCTGCGCCGATCTTTAAAAATCAAAAAGGCGGACTTTAAAGATAA	32060
Qy	381	TGGGGAGTACTGTCGTGTGTCGGCAGGAGAGACCTCGGTATGCTCACCGTCAGGCG	440
Db	32061	AGGCGAATATGTGTGACTGTGGCAGACAGACCAAGGCAATGTTACTGTGTGAGGC	32120
Qy	441	TCTACCCATCAAGTTTCAAGAGGTTCTGAGAACGAAGAGGGCCACAGAGGGGCAACAGC	500
Db	32121	TCGACTAATAAAAGTGGMAAGCCCTCTGTACGGAGTAGAGGTGTTTGTGGTGAACACAGC	32180
Qy	501	CGTGCTCGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAGTGGTGGAGGGGCGATG---A	557
Db	32181	CCACTTTGAAATTGAACTTTCTGAACCTGATTTTCACGCCAGTGGGAAGTGAAGGACA	32240
Qy	558	GACCTCAGAGATGGAGACAGACACAGCCTGAGGCAGGACGGGCCACAGTGTGAGCTGCA	617
Db	32241	GCCTTTGACAGCTTCCCTGACTGTGAATCATTTGAGGATGAAGAAGACATATTTCTGAT	32300
Qy	618	GATCCGGCGCTCGTGGCAGAGGACGCTGGGAGTACCTGTGCATGTGCGGGAAGGAGAG	677
Db	32301	CCTTCAATACGTGCAGCTGGGTATGACAGAGAGGTTTCTTCCAGGCTGCTAATGCCAA	32360
Qy	678	GACCTGAGCCATGCTCACCGTCAGGGCCATGCTTCCAAGTTTCATAGAGGTCTGAGGAA	737
Db	32361	ATCTGACGCCAATCTCAAAAGTGAAGAAATTGCCCTCTTATCTTTCATCACACCTCTCAGTGA	32420
Qy	738	TGAAGAGGCCACAGAAAGGGGACACGGCCACCGTGTGGTGTGAGCTGAGCAAGGGCGC---	794
Db	32421	TGTTAAAGTCTTCGAAAGATGAGGCTAAGTTTGAGTGTGAAGTATCCAGGGAGGCCAA	32480
Qy	795	ACCGTGGAGTGGAGAAAGGGCATGAGACCTCAGAGATGGGACAGACACAGCTGAG	854
Db	32481	AACATTTCCGTTGGCTTAAAGGAAACCCAGGAAATCACAGTGTATGACAGATTTGAGCTTAT	32540
Qy	855	GCAGGACGGGTCCAGGTGTGAGCTGCAGATCCGTGGCCTGGCTGTGGTGTGATGCCGGGA	914
Db	32541	AAAGGATGGCACTTAAGCATTTCAATGGTGTATCAAGTCAGCTGCTTTTGAAGATGAAGCAAA	32600
Qy	915	GTACTCGTGTGTGTCGGGCAGGAGGACCTCAGCCACACTCACTGTCAAGGCCCTTGCC	974
Db	32601	ATACATGTTTGAAGCTGAAGATAAGCACAAAGTGGCAAACTGATCATTTGAAGGATCCG	32660
Qy	975	TGCCAGATTCAATAGAGATGTGAAAAA	1001
Db	32661	GCTCAAATTTCTCACCCCTCTCAAAGA	32687

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:02:07 ; Search time 175.678 Seconds
(without alignments)
9323.380 Million cell updates/sec

Title: US-10-077-130-4_COPY_3120_4120
Perfect score: 1001
Sequence: 1 aaagatgggaagctgag.....tcagagcccaaggcggtgtt 1001

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	248.4	24.8	2768	4	US-09-774-528-426
2	69.6	7.0	7218	1	Sequence 426, Appl
3	55.4	5.5	3766	4	Sequence 14, Appl
4	55.4	5.5	3778	4	Sequence 1651, Ap
5	52.6	5.3	1926	3	Sequence 359, App
6	52.6	5.3	1931	2	Sequence 4, Appli
7	49.6	5.0	5207	4	Sequence 2, Appli
8	49.6	5.0	5207	4	Sequence 1, Appli
9	49.6	5.0	5207	4	Sequence 3, Appli
10	49.6	5.0	5207	4	Sequence 1, Appli
11	49.6	5.0	5207	4	Sequence 3, Appli
12	49.2	4.9	1505	1	Sequence 1, Appli
13	47.2	4.7	3575	4	Sequence 1395, Ap
14	47	4.7	18955	4	Sequence 13343, A
15	47	4.7	30678	4	Sequence 12818, A
16	46.8	4.7	1926	3	Sequence 2, Appli
17	46.8	4.7	1926	3	Sequence 3, Appli
18	46.8	4.7	2580	3	Sequence 2, Appli
19	46.8	4.7	2580	3	Sequence 2, Appli
20	46.8	4.7	5452	2	Sequence 1, Appli
21	46.8	4.7	8705	4	Sequence 14, Appli
22	46.8	4.7	9600	3	Sequence 1, Appli
23	46.8	4.7	9600	3	Sequence 1, Appli
24	46.8	4.7	10596	1	Sequence 15, Appl
25	46.8	4.7	10596	1	Sequence 15, Appl
26	46.8	4.7	10596	1	Sequence 15, Appl
27	46.8	4.7	10596	1	Sequence 15, Appl

28 46.8 4.7 10596 2 US-08-194-087-15 Sequence 15, Appl
29 46.8 4.7 10596 5 PCT-US93-04648-15 Sequence 15, Appl
30 46.8 4.7 16080 4 US-09-724-566A-48 Sequence 48, Appl
31 46.8 4.7 16080 4 US-09-471-669A-48 Sequence 48, Appl
32 46.4 4.6 2949 3 US-09-412-554A-3 Sequence 3, Appli
33 46.2 4.6 2225 4 US-09-620-312D-1046 Sequence 1046, Ap
34 45.6 4.6 1701 4 US-09-902-540-7774 Sequence 7774, Ap
35 45.6 4.6 6250 4 US-09-902-540-773 Sequence 773, App
36 44.8 4.5 809 4 US-09-949-016-2543 Sequence 2543, Ap
37 44.6 4.5 1257 4 US-09-826-509-572 Sequence 572, App
38 44.6 4.5 1296 1 US-07-816-283-9 Sequence 9, Appli
39 44.6 4.5 1296 1 US-08-417-103-9 Sequence 1321, Ap
40 44.6 4.5 1413 4 US-09-016-434-1321 Sequence 14385, A
41 44 4.4 8260 4 US-09-949-016-14285 Sequence 120, App
42 43.2 4.3 6202 4 US-09-774-528-120 Sequence 13134, A
43 43.2 4.3 31467 4 US-09-949-016-13134 Sequence 11907, A
44 43.2 4.3 31868 4 US-09-949-016-11907 Sequence 12147, A
45 43.2 4.3 767677 4 US-09-949-016-12147 Sequence 12147, A

ALIGNMENTS

RESULT 1
US-09-774-528-426
; Sequence 426, Application US/09774528
; Patent No. 6743619
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Yang, Yonghong
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6743619el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802
; CURRENT APPLICATION NUMBER: US/09/774,528
; CURRENT FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt_Fl_genes Version 2.0
; SEQ ID NO 426
; LENGTH: 2768
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72)..(2738)
US-09-774-528-426

Query Match 24.8%; Score 248.4; DB 4; Length 2768;
Best Local Similarity 57.4%; Pred. No. 3.8e-54;
Matches 510; Conservative 0; Mismatches 366; Indels 12; Gaps 3;
QY 75 GCACAGCGCAGGCAAGCGGAGTACCGGGGACTACAGCTGCGAGGCCAGGGGCCAGAGGGT 134
DB 1874 GGAGGTGCGCCCGCCAGCTGCACGGGGCGCAGCTGAAGTTCGTGGCCACCGCATTTGAGAG 1933
QY 135 CTCCTTCGCGCTGCACATCACAGACCCCAAGATGATGTTTGCAGAGGACGAGTCAAGTCA 194
DB 1934 CAGCATCGGATGAGGTTCGCGGGCGGCCCGCCAGGCTGACTGCCCAACAGCCGCGCCG 1993
QY 195 TAAT-----CAGGTGCGAGGCTGAGGGGGGGCCAGTGCCTGCTGTGAGCTGTGAGGTGCG 248
DB 1994 AGTTCGCGGGAGGTGTGCTGCGCTCGGCTGCACAGAGGCGGAGCTGCTGGCTGAGCTGTC 2053

QY 249 CCAGGCCAGACGAGGTGACGTGGTACAAAGATGGGAAGAGCTGAGCTCCAGCTCAA 308
Db 2054 AGATCAGGCTGGCTGTGACGTGGTGAAGATGTCACACACTGTCCCGAGGCCCA 2113
QY 309 AGTGGGCATGAGAGTCAAAAGGTGCACACGAGAGTGGTGTGTCACAGCGGGCAAGC 368
Db 2114 GTATGAGGTGACAGCATCGCGCGGGCGGGTGTCTCTGTGCGAGATGTGGCCGGGA 2173
QY 369 AGATGCTGGGAGTACAGCTGTGAGGCTGGGGCCAGAGAGTCTCTTCCACTGCACAT 428
Db 2174 CGATGAGCCCTCTACGATGGCTGACCGCGGGGCCGATCGCTACAGCTCTCCGT 2233
QY 429 CACAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATAATAGGTGCAGGCTGA 488
Db 2234 GCAAGGCTCGCGCTTTCTGCACAGGACATGGCGGCGACTGT--GTGATGCCGT 2290
QY 489 GCGGGGACACTGCATCTGAGCTGTGAGTGGGCCAGCCCGCCAGAGGAGTCACTG 548
Db 2291 GGTGGGGCGGGCGAGTTTGAAGTGTGAGACCTCCGAAGCCACGTCACGTCGACTG 2350
QY 549 GTACAAGGACGGAAGAGCTGAGTCCAGCTCAAGTCAAGTACGATGGAGGTCAAGGGCTG 608
Db 2351 GTACAAGATGCATGGAGCTGGGCACCTCCGCTGAGCGCTTCTTGCAGAGGATGTGGG 2410
QY 609 CACAGC--AAGGCTGTAGTGCAGAGTGGGCAAGCAGATGCTGGGGAGTACAGCTG 665
Db 2411 GACGGGACCGCTGGTGGCAGCCACAGTCAAGGAGGATGAAGCACCCTACTCTG 2470
QY 666 CGAGCTGGGGCCAGAGTCTCTTCACTGACATCAAGAGCCCAAGGAGTGT 725
Db 2471 CGCGTGGCGAGGACTCTGTGGACTTCGGCTCCGCTCTGTGAGCCCAAGGCGGTGT 2530
QY 726 TGCCAAGGACAGTGTGGTGCATAATGAGTGGGACTGAGGCGAGGCGCAGTCCACACT 785
Db 2531 TGCCAAGGACAGCGCGGTGCAGGAGGTGAGCTGAGTGGGGGCGAGCCACGCT 2590
QY 786 GAGCTGTGAGTGGCCAGGCGCCAGACAGAGGTGACGTGGTACAGGATGGGAAGACT 845
Db 2591 GAGCTGTGAGTGGCCAGGACAGATGAGGTGACGTGGTACAGGACGGAAGT 2650
QY 846 GAGCTCCAGTTCGAAGTGGCATAGAGCTCGGGCTGCATGGCGAGCTGGTGTGCA 905
Db 2651 GAGCTCCAGTTCGAAGTGGCATGAGGCTGTGGGCTGTATGCGAGGCTGGTGTGCA 2710
QY 906 CGAGCAGCCAGGAGATGCTGGGGAGTACACCTGTGAGGTGGGG 953
Db 2711 GCAGTGGGCCAGGACAGCTCCGGATAGTACAGCTGTGAAGCCAGGTG 2758

RESULT 2

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367

GENERAL INFORMATION:

; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match 7.0%; Score 69.6; DB 1; Length 7218;

Best Local Similarity 4.5%; Pred.No.5.2e-08;
Matches 18; Conservative 233; Mismatches 147; Indels 0; Gaps 0;

QY 28 TTGAAGTGCATGTAGAGGCCAAAGGCTGCAGACGGAGGCTGTGTGTCAGCAGCAGGC 87
Db 1441 TTGCTACRR 1382
QY 88 AAGCAGATCGCGGGACTACAGCTGCAGACGGGCGCAGAGGCTCTCTCCCGCTG 147
Db 1381 RRR 1322
QY 148 CACATCAGACGCCAAGATGATGTTTCAAGAGCAGTCAGTGCATATAGGTGTCAG 207
Db 1321 RRR 1262
QY 208 GCTGAGCGGGCCAGTGCATGCTGAGTGTGAGTGGCCAGGCCAGCAGAGGTG 267
Db 1261 RRR 1202
QY 268 ACCTGGTACAGGATGGGAAGAGTGCAGCTCCAGCTCAAAAGTGGCATGGAGTCAA 327
Db 1201 RRR 1142
QY 328 GGGTCACACGAGGCTGTGCTGCCACAGCGGCAAGCAGATGCTGGGAGTACAGC 387
Db 1141 RRR 1082
QY 388 TGTGAGCTGGGGCCAGAGATCTCTTCCACTGCA 425
Db 1081 RRR 1044

RESULT 3

US-09-949-016-1651
; Sequence 1651, Application US/09949016
; Patent No. 6812339

GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

[illegible]

```

RESULT 7
US-09-858-664A-1
; Sequence 1, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-858-664A-1

Query Match          5.0%; Score 49.6; DB 4; Length 5207;
Best Local Similarity 57.9%; Pred. No. 0.0065;
Matches 88; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy      520 GTGGCCACGCCAGACGGAGTGACGTGGTACAGAGCGGAGAGCTGACGCTCCAGC 579
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      340 GAGGCGGACCCACAGCCCTCGGTGACCTGGTACAGGACAGCGCTGGTGGACAGC 339
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      580 TCAAAAGTAGCATGAGAGTCAAGGGCTGCACACGAAAGCTCGTAGTCGACGAGGTGGGC 639
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      400 ACCCGGCTTAGCGAGCAGCAAGAGGACCAATACCTCCCTGGTCTGAGGCAATGTGGCC 459
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```



```
QY 640 AAAGCAGATGCTGGGGAGTACAGCTGCGAGGC 671
Db 460 TCGAAGGATGCGCGGCTTTACACCTGCTGGC 491

RESULT 8
US-10-274-978-1
; Sequence 1, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
US-10-274-978-1

Query Match 5.0%; Score 49.6; DB 4; Length 5207;
Best Local Similarity 57.9%; Pred. No. 0.0065;
Matches 88; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 520 GTGGCCAGCCCCCAGCAGGTGACGTGCGAGGAGCGGGAAGAGTGTGAGCTCCAGC 579
Db 340 GAGGCGACCCACACAGCCCTCGTGACCTGGTACAGGACAGCGTCCAGCTGTGGACAGC 399

QY 580 TCAAAAGTACGATGAGGTCAAGGCTGCACAGAGCGTGTAGTGCAGCAGGTGGGC 639
Db 400 ACCCGGCTTAGCCAGCAGCAAGAGGACCACATACCTCCCTGGTGTGAGGCATGTGGCC 459

RESULT 9
US-10-274-978-3
; Sequence 3, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
US-10-274-978-3

Query Match 5.0%; Score 49.6; DB 4; Length 5207;
Best Local Similarity 57.9%; Pred. No. 0.0065;
Matches 88; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 640 AAAGCAGATGCTGGGGAGTACAGCTGCGAGGC 671
Db 460 TCGAAGGATGCGCGGCTTTACACCTGCTGGC 491

RESULT 10
US-10-697-263-1
; Sequence 1, Application US/10697263
; Patent No. 6812014
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV2
; CURRENT APPLICATION NUMBER: US/10/697,263
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-697-263-1

Query Match 5.0%; Score 49.6; DB 4; Length 5207;
Best Local Similarity 57.9%; Pred. No. 0.0065;
Matches 88; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 520 GTGGCCAGCCCCCAGCAGGTGACGTGCGAGGAGCGGGAAGAGTGTGAGCTCCAGC 579
Db 340 GAGGCGACCCACACAGCCCTCGTGACCTGGTACAGGACAGCGTCCAGCTGTGGACAGC 399

QY 580 TCAAAAGTACGATGAGGTCAAGGCTGCACAGAGCGTGTAGTGCAGCAGGTGGGC 639
Db 400 ACCCGGCTTAGCCAGCAGCAAGAGGACCACATACCTCCCTGGTGTGAGGCATGTGGCC 459

RESULT 11
US-10-697-263-3
; Sequence 3, Application US/10697263
; Patent No. 6812014
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV2
; CURRENT APPLICATION NUMBER: US/10/697,263
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
```


CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1395:
SEQUENCE CHARACTERISTICS:
LENGTH: 3575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 9402646
US-09-023-655-1395

Query Match 4.7%; Score 47.2; DB 4; Length 3575;
Best Local Similarity 54.7%; Pred. No. 0.024; 78; Indels 0; Gaps 0;
Matches 94; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 785 TGAGCTGTGAGTGGCCAGCCAGAGAGTGCAGTGTGACAGAGTGGTACAAAGGATGGGAAGC 844
DB 1392 TCAAGTCGAGGTGCTGATGAGAAAGTGACGGCGCAAGTGTATGAAGATGGGTGCGAGG 1451

QY 845 TGAGCTCAGTTGAAAGTGCATAGAGGCTGCGGGTGCATGCGGAGCTGTGTGTCG 904
DB 1452 TCGCGCCAGCAAGAGGATCACCATTTCCCATGTAGCAGGTTCACAAAGCTGTGTATCG 1511

QY 905 AGCAGCGAGCCAGCAGATGCTGGGAGTACACCTGTGAGCTGCGGGCCA 956
DB 1512 ATGAGCTCGCCCGGAGGATGAGGAGACTACACGTTTGTGCTGACGGCTA 1563

RESULT 14
US-09-949-016-13343
Sequence 13343, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13343
LENGTH: 18955
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(18955)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13343

Query Match 4.7%; Score 47; DB 4; Length 18955;
Best Local Similarity 44.6%; Pred. No. 0.048;
Matches 226; Conservative 0; Mismatches 280; Indels 1; Gaps 1;

QY 179 AGGAGCAGTCAGTCATTAATGAGGTGACAGGCTGAGCGGGGCCAGTGCCTATGCTGAGCT 238
DB 3364 AGGAGGGGGAGGGGGAAGAGGAGGAGGGGGGGAAGAGCGGCAAGAGAGGTTGAGCGGG 3423

QY 239 GTGAGGTGCCCCAGGCCCCAGACGAGGTGACCTGTGTACAAAGGATGGGAAGAGCTGAGCT 298
DB 3424 AGGAGGAGGGGAGGAGGAGGGGGGAGGAGTAAGAGGTGGNAGGGGAGCGGAAGGGGAGA 3483

QY 299 CCAGCTCAAAAGTGGGCATATGAGGTCAAAAGGTGCAACGAGGCTGTGTGCTGCCACAGG 358
DB 3484 GTAGGAGGAGGAGGAGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3543

QY 359 CGGCAAGACAGATGCTCGGGGAGTACAGCTGTGAGGCTGGGGGCCAGAGAGTCTCTTCC 418
DB 3544 AGGGGAAGGGAGAGGAGGGGGGAGAAAGAGGGAGGAGTAGGAGGGAGCAGAAAGGGG 3603

QY 419 ACCTGCACATCACAGAGCCCAAGGGGGTGTGTTTGGGAAGGAGCAGTCAAGTCAATAGAGG 478
DB 3604 AGGAATAGGAGGGGAGGAGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3663

QY 479 TGCAGGCTGAGCGGGGACCACTGCTCATGCTGTGAGGTGTGAGGTGCCCCAGCCCCAGACGG 538
DB 3664 GGGAGGAGTAGGAGGTGAGCA-GAAGGGGAGGAGTAGGAGGGGAGGAGGAGGAGGAGGAGG 3722

QY 539 AGGTGACGTGTACAGGACGCGGAGAGCTGAGCTCCAGCTCAAAAGTACGTCATGAGG 598
DB 3723 AGAGGAGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 3782

QY 599 TCAAGGGCTGCACACAGCAAGGCTGTGTAGTGCAGCAGCTGGGCAAAAGCAGATGCTGGGGAGT 658
DB 3783 AGGAGGAGGAGCAGAAAGGAGGAGTGTAGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGG 3842

QY 659 ACAGTCGAGGCTGGGGGCCAGAGAG 685
DB 3843 GAGGAGAAAGGGAGGAGTAGGAGGAGG 3869

RESULT 15
US-09-949-016-12818
Sequence 12818, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12818
LENGTH: 30678
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(30678)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12818

Query Match 4.7%; Score 47; DB 4; Length 30678;
Best Local Similarity 44.6%; Pred. No. 0.057;
Matches 226; Conservative 0; Mismatches 280; Indels 1; Gaps 1;

QY 179 AGGAGCAGTCAGTCATTAATGAGGTGACAGGCTGAGCGGGGCCAGTGCCTATGCTGAGCT 238
DB 15090 AGGAGGGGAGGGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 15149

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:32:32 ; Search time 627.052 Seconds
(without alignments)
9503.485 Million cell updates/sec

Title: US-10-077-130-4_COPY_3120_4120
Perfect score: 1001
Sequence: 1 aaagatgggaagagctgag.....tcagagcccaaggcggtgtt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5544816 seqs, 2976611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	23907	13	US-10-077-130-6
2	1001	100.0	24120	13	US-10-077-130-4
3	965	96.4	20489	18	US-10-723-860-5701
4	248.4	24.8	2768	17	US-10-120-988-426
5	185.6	18.5	578	13	US-10-027-632-272054
6	185.6	18.5	578	17	US-10-027-632-272054
7	123.2	12.3	1645	9	US-09-726-643-22
8	123.2	12.3	1645	9	US-10-042-141-22
9	123.2	12.3	1645	19	US-10-919-272-22
10	123.2	12.3	2170	17	US-10-264-049-850
11	97.4	9.7	14061	17	US-10-093-463-73

12	97.4	9.7	14109	17	US-10-093-463-71	Sequence 71, Appl
13	95.8	9.6	2534	17	US-10-108-260A-1946	Sequence 1946, Ap
14	95.8	9.6	7564	18	US-10-476-397-15	Sequence 15, Appl
15	88.4	8.8	2254	9	US-09-726-643-44	Sequence 44, Appl
16	88.4	8.8	2254	13	US-10-042-141-44	Sequence 44, Appl
17	88.4	8.8	2254	19	US-10-919-272-44	Sequence 44, Appl
18	68	6.8	716	10	US-09-822-846-117	Sequence 117, App
19	66.8	6.7	9591	10	US-09-764-891-6043	Sequence 6043, Ap
20	66.8	6.7	9591	14	US-10-091-438-268	Sequence 268, App
21	66.8	6.7	12415	10	US-09-764-891-6044	Sequence 6044, Ap
22	66.8	6.7	12415	14	US-10-091-438-269	Sequence 269, App
23	66.8	6.7	20565	10	US-09-764-891-6045	Sequence 6045, Ap
24	66.8	6.7	20565	14	US-10-091-438-270	Sequence 270, App
c	64.8	6.5	400	17	US-10-242-535A-8373	Sequence 8373, Ap
25	64.8	6.5	400	17	US-10-085-783A-8373	Sequence 8373, Ap
26	64.8	6.5	400	17	US-10-085-783A-8373	Sequence 8373, Ap
27	62	6.2	81940	9	US-09-759-508B-1	Sequence 1, Appl
28	62	6.2	81940	10	US-09-960-706-1092	Sequence 1092, Ap
29	62	6.2	81940	10	US-09-873-319-724	Sequence 724, App
30	62	6.2	81940	18	US-10-723-860-132	Sequence 132, App
31	62	6.2	81940	18	US-10-656-873A-1	Sequence 1, Appl
32	62	6.2	82027	19	US-10-278-698-1034	Sequence 1034, Ap
33	62	6.2	82027	19	US-10-278-698-1045	Sequence 1045, Ap
34	61.6	6.2	1266	10	US-09-822-846-116	Sequence 116, App
c	61.4	6.1	2884	17	US-10-094-749-1001	Sequence 1001, Ap
35	55.4	5.5	3778	18	US-10-723-860-2224	Sequence 2224, Ap
36	55.4	5.5	3966	18	US-10-357-930-21853	Sequence 21853, A
37	55.4	5.5	3966	18	US-10-357-930-25171	Sequence 25171, A
38	55.4	5.5	3966	18	US-10-357-930-27704	Sequence 27704, A
39	55.4	5.5	3966	18	US-10-723-860-6456	Sequence 6456, Ap
40	54.2	5.4	2029	18	US-10-287-092-11	Sequence 11, Appl
41	53.8	5.4	3535	17	US-10-287-092-13	Sequence 13, Appl
42	53.8	5.4	3673	17	US-10-723-860-4453	Sequence 4453, Ap
43	52.6	5.3	3613	18	US-10-723-860-8268	Sequence 8268, Ap
44	52.6	5.3	4176	18	US-10-723-860-8268	Sequence 8268, Ap
45	49.6	5.0	3935	17	US-10-108-260A-534	Sequence 534, App

ALIGNMENTS

RESULT 1
US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCF1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match 100.0%; Score 1001; DB 13; Length 23907;
Best Local Similarity 100.0%; Pred. No. 7.1e-266;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAAGATGGGAAGAGCTGAGCTCCAGCTTGAAGCTGATGTAGAGCCAAAGGCTGCAGA 60
Db 3049 AAAGATGGGAAGAGCTGAGCTCCAGCTTGAAGCTGATGTAGAGCCAAAGGCTGCAGA 3108
Qy 61 CGGAGGCTGGTGGTGCAGCAGCAGGAGCAAGACGGATGCCGGGACTACAGCTGCCAGGCC 120
Db 3109 CGGAGGCTGGTGGTGCAGCAGCAGGAGCAAGACGGATGCCGGGACTACAGCTGCCAGGCC 3168

```
QY 121 AGGGCCAGAGGGTCTCTTCCGCTGCACATCACAGAGCCCAAGATGATGTTTCAAG 180
Db |||||
QY 3169 AGGGCCAGAGGGTCTCTTCCGCTGCACATCACAGAGCCCAAGATGATGTTTCAAG 3228
Db |||||
QY 181 GAGCAGTCAGTCATAATAGAGTGCAGCTGAGGGCGGGCCAGTGCATGCTGAGCTGT 240
Db |||||
QY 3229 GAGCAGTCAGTCATAATAGAGTGCAGCTGAGGGCGGGCCAGTGCATGCTGAGCTGT 3288
QY 241 GAGGTGGCCAGGCCAGACGAGAGTGCAGTGGTACAAGATGGGAAGAGCTGAGCTCC 300
Db |||||
QY 3289 GAGGTGGCCAGGCCAGACGAGAGTGCAGTGGTACAAGATGGGAAGAGCTGAGCTCC 3348
QY 301 AGCTCAAAAGTGGGATGAGAGTCAAAAGGTGCACACGAGAGTGTGCTGTCACAGCG 360
Db |||||
QY 3349 AGCTCAAAAGTGGGATGAGAGTCAAAAGGTGCACACGAGAGTGTGCTGTCACAGCG 3408
QY 361 GCGAAGCAGATGCTGGGAGTACAGCTGTGAGGCTGGGGCCAGAGAGTCTCTTCCAC 420
Db |||||
QY 3409 GCGAAGCAGATGCTGGGAGTACAGCTGTGAGGCTGGGGCCAGAGAGTCTCTTCCAC 3468
QY 421 CTGCACATCACAGAGCCCAAGGGGTGTTTGGGAAGGAGCAGTCAAGTGCATAATGAGGTG 480
Db |||||
QY 3469 CTGCACATCACAGAGCCCAAGGGGTGTTTGGGAAGGAGCAGTCAAGTGCATAATGAGGTG 3528
QY 481 CAGGCTGAGGGGGGACCACTGCCATGTGAGCTGTGAGGTGGGCCAGCCGCCAGACGAG 540
Db |||||
QY 3529 CAGGCTGAGGGGGGACCACTGCCATGTGAGCTGTGAGGTGGGCCAGCCGCCAGACGAG 3588
QY 541 GTGAGCTGTACAAGGAGCGGAAGAGTGCAGTCCAGCTCAAAAGTACGATGAGGTGTC 600
Db |||||
QY 3589 GTGAGCTGTACAAGGAGCGGAAGAGTGCAGTCCAGCTCAAAAGTACGATGAGGTGTC 3648
QY 601 AAGGCTGCACACGAAGGCTGTGTAGTCAGAGGTGGGCAAGAGCAGATGCTGGGGAGTAC 660
Db |||||
QY 3649 AAGGCTGCACACGAAGGCTGTGTAGTCAGAGGTGGGCAAGAGCAGATGCTGGGGAGTAC 3708
QY 661 AGCTGCGAGGTGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGCA 720
Db |||||
QY 3709 AGCTGCGAGGTGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGCA 3768
QY 721 GTGTTTGCAGAGGAGCAGTGGTGCATAATGAGGTGGGAGTGCAGGAGGGCCAGTGC 780
Db |||||
QY 3769 GTGTTTGCAGAGGAGCAGTGGTGCATAATGAGGTGGGAGTGCAGGAGGGCCAGTGC 3828
QY 781 ACATGAGCTGTGAGGTGGCCAGGCCAGACAGAGTGCAGTGGTACAAGATGGGAAG 840
Db |||||
QY 3829 ACATGAGCTGTGAGGTGGCCAGGCCAGGCCAGACAGAGTGCAGTGGTACAAGATGGGAAG 3888
QY 841 AAGCTGAGCTCCAGTTGGAAGTGGCATAGAGGCTGGGGCTGCATGCGGCAGCTGGTG 900
Db |||||
QY 3889 AAGCTGAGCTCCAGTTGGAAGTGGCATAGAGGCTGGGGCTGCATGCGGCAGCTGGTG 3948
QY 901 GTGCAGAGGAGGAGGAGAGTCTGGGAGTACACCTGTGAGGCTGGGGCCAGCGG 960
Db |||||
QY 3949 GTGCAGAGGAGGAGGAGAGTCTGGGAGTACACCTGTGAGGCTGGGGCCAGCGG 4008
QY 961 CTCTCTCTCCACCTGGATGTTTCAAGAGCCCAAGCGGTGTT 1001
Db |||||
QY 4009 CTCTCTCTCCACCTGGATGTTTCAAGAGCCCAAGCGGTGTT 4049
```

RESULT 2

```
US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Accon, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Thereof
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
```

```
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-4
```

```
Query Match 100.0%; Score 1001; DB 13; Length 24120;
Best Local Similarity 100.0%; Pred. No. 7,1e-266;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGATCGGAGAGCTGAGCTCCAGCTTGAAGCTGCATGTAGAGGCCAAAGGCTGCAGA 60
Db |||||
QY 3120 AAAGATCGGAGAGCTGAGCTCCAGCTTGAAGCTGCATGTAGAGGCCAAAGGCTGCAGA 3179
QY 61 CGGAGGCTGTGTGTCAGCAGGCGAGCAAGACGGATGCCGGGACTACAGCTGCGAGGCC 120
Db |||||
QY 3180 CGGAGGCTGTGTGTCAGCAGGCGAGCAAGACGGATGCCGGGACTACAGCTGCGAGGCC 3239
QY 121 AGGGGCCAGAGGGTCTCTTCCGCTGCACATCACAGAGCCCAAGATGATGTTTGCAGAG 180
Db |||||
QY 3240 AGGGGCCAGAGGGTCTCTTCCGCTGCACATCACAGAGCCCAAGATGATGTTTGCAGAG 3299
QY 181 GAGCAGTCAGTCATAATGAGGTGCAGGTCGAGCGGGGGGCCAGTGCATGCTGAGCTGT 240
Db |||||
QY 3300 GAGCAGTCAGTCATAATGAGGTGCAGGTCGAGCGGGGGGCCAGTGCATGCTGAGCTGT 3359
QY 241 GAGGTGCCCCAGGCCAGAGCGAGGTGACGTGGTACAGAGATGGGAAGAGCTGAGCTCC 300
Db |||||
QY 3360 GAGGTGCCCCAGGCCAGAGCGAGGTGACGTGGTACAGAGATGGGAAGAGCTGAGCTCC 3419
QY 301 AGCTCAAAAGTGGGATGAGAGGTCAAAAGGTGCACACGAGAGGTGCTGCTGCCACAGCG 360
Db |||||
QY 3420 AGCTCAAAAGTGGGATGAGAGGTCAAAAGGTGCACACGAGAGGTGCTGCTGCCACAGCG 3479
QY 361 GCGAAGCAGATGCTGGGAGTACAGCTGTGAGCTGGGGCCAGAGAGTCTCTTCCAC 420
Db |||||
QY 3480 GCGAAGCAGATGCTGGGAGTACAGCTGTGAGCTGGGGCCAGAGAGTCTCTTCCAC 3539
QY 421 CTGCACATCACAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATAATGAGGTG 480
Db |||||
QY 3540 CTGCACATCACAGAGCCCAAGGGGTGTTTGCAGAGGAGCAGTCAAGTGCATAATGAGGTG 3599
QY 481 CAGGCTGAGCGGGGAGCACTGCCATGCTGAGCTGTGAGGTGGGCCAGAGCCCAAGAGAG 540
Db |||||
QY 3600 CAGGCTGAGCGGGGAGCACTGCCATGCTGAGCTGTGAGGTGGGCCAGAGCCCAAGAGAG 3659
QY 541 GTGACGTGGTACAAAGGAGCGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGAGTGC 600
Db |||||
QY 3660 GTGACGTGGTACAAAGGAGCGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGAGTGC 3719
QY 601 AAGGGCTGCACACGAAGGCTGGTAGTGCAGAGGTGGGCAAGAGCAGATGCTGGGGAGTAC 660
Db |||||
QY 3720 AAGGGCTGCACACGAAGGCTGGTAGTGCAGAGGTGGGCAAGAGCAGATGCTGGGGAGTAC 3779
QY 661 AGCTGCGAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGCA 720
Db |||||
QY 3780 AGCTGCGAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGCA 3839
QY 721 GTGTTTGCAGAGGAGCAGTGGTGCATAATGAGGTGGGAGTGCAGGAGGGCCAGTGC 780
Db |||||
QY 3840 GTGTTTGCAGAGGAGCAGTGGTGCATAATGAGGTGGGAGTGCAGGAGGGCCAGTGC 3899
```


Db 3480 GGCAAGCAGATGCTGGGAGTACAGCTGTGAGGCTGGGGCCAGAGAGTCTCTCTTCCAC 3539
Qy 421 CTGCACATCAGAGCCCAAGGGGTGTTTCGAAGGAGCAGTCAAGTGCATAAATGAGGTG 480
Db 3540 CTGCACATCAGAGCCCAAGGGGTGTTTCGAAGGAGCAGTCAAGTGCATAAATGAGGTG 3599
Qy 481 CAGGCTGAGGGGGACCACTGCCATGCTGAGCTGTGAGGTGGCCACGCCCCAGACGGAG 540
Db 3600 CAGGCTGAGGGGGACCACTGCCATGCTGAGCTGTGAGGTGGCCACGCCCCAGACGGAG 3659
Qy 541 GTGAGCTGGTACAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTG 600
Db 3660 GTGAGCTGGTACAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTG 3719
Qy 601 AAGGCTGCACACGAAGGCTGTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGGAGTAC 660
Db 3720 AAGGCTGCACACGAAGGCTGTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGGAGTAC 3779
Qy 661 AGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTCAACTGSCATCACAGAGCCCAAGGCA 720
Db 3780 AGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTCAACTGSCATCACAGAGCCCAAGGCA 3839
Qy 721 GTGTTTGCAGAGAGCAGTGTGGTGCATAATGAGGTGGGACTGAGGACGGGCCAGTGCC 780
Db 3840 GTGTTTGCAGAGAGCAGTGTGGTGCATAATGAGGTGGGACTGAGGACGGGCCAGTGCC 3899
Qy 781 ACACGTAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGTGTAAGGATGGGAAG 840
Db 3900 ACACGTAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGTGTAAGGATGGGAAG 3959
Qy 841 AAGCTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTGCGGCTGCATCGGCAGCTGGTG 900
Db 3960 AAGCTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTGCGGCTGCATCGGCAGCTGGTG 4019
Qy 901 GTGCAGCAGGACGGCAGGACAGTCTGGGAGTACACTGTGAGGCTGGGGGCCAGCGG 960
Db 4020 GTGCAGCAGGACGGCAGGACAGTCTGGGAGTACACTGTGAGGCTGGGGGCCAGCGG 4079
Qy 961 CTCTCTTCCACCTGGATGTTTCAGAGCCCAAGGGGTGTT 1001
Db 4080 CTCTCTTCCACCTGGATGTTTCAGAGCCCAAGGGGTGTT 4120

RESULT 4

US-10-988-426
; Sequence 426, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyan
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 426
; LENGTH: 2768
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72)..(2738)
US-10-120-988-426

Query Match

24.8%; Score 248.4; DB 17; Length 2768;

Best Local Similarity 57.4%; Pred. No. 1.8e-58;
Matches 510; Conservative 0; Mismatches 366; Indels 12; Gaps 3;
Qy 75 GCACAGCGCAGGCAAGACGGATGCCGGGACTACAGCTGCGAGGCCAGGGCCAGAGGTT 134
Db 1874 GGAGGTGCCGCCAGCCTGACCGGGGCGACCTGAAAGTTCGTGGCCAAACGCAATTGAGAG 1933
Qy 135 CTCCTTCGCGCTGCACATCACAGAGCCCAAGATGATGTTTCAAAAGGAGCAGTCAAGTCA 194
Db 1934 CAGCATCCGATGAGGTCCGGGCGGCCAGGGCTGACTGCCAACAAAGCCGACGCCGC 1993
Qy 195 TAAT-----GAGGTGACGCTGAGGGCGGGGCCAGTGCCTATGCTGAGCTGTGAGGTGGC 248
Db 1994 AGCTGCCCGGAGGTGCTGCTCGGCTGCACAGAGAGGCGAGCTGTGGGTGAGCTGTC 2053
Qy 249 CCAGGCCCAGACGAGGTGAGCTGATCAAGATGGAAGAGCTGAGCTCCAGCTCAAA 308
Db 2054 AGATCAGGCTGCGGCTGTGAGCTGCTGAAGATGTTGCGACACTGTCCCCAGGCCCCAA 2113
Qy 309 AGTGGCATGAGGTCAAAAGGGTGCACACGAGGCTGTGTGTCACAGGGCGGCAAAAGC 368
Db 2114 GTATGAGGTGACGATCGCGCGGGCGGGTCTCTTTGTCGAGATGTGGCCCGGA 2173
Qy 369 AGATGCTGGGAGTACAGCTGTGAGGTGGGGGCCAGAGTCTCTTCCACCTGCAAT 428
Db 2174 CGATGCAGGCTCTTACAGGTGCTGAGCGCGGGGGCGCATCGCTACCAGCTCTCGT 2233
Qy 429 CACAGAGCCCAAGGGGTGTTTGAAGAGCAGTCAAGTGCATAATGAGTGCAGGTGA 488
Db 2234 GCAAGGCTCTCGCGCTTTTTCGCAAGAGCATGGCGGCGAGCTGT---GTGGATGCCGT 2290
Qy 489 GCGCGGGACCACTGCCATGTGAGCTGTGAGGTGGCCAGCCAGCCAGAGGTCAGCTG 548
Db 2291 GGCTGGGGGCCGCGCAGTTTGAAGTGTGAGACCTCCGAAGCCCAAGTCCACGTGCACTG 2350
Qy 549 GTACAAGGACGGGAAGAGTGTGAGCTCAGCTCAAAAGTACGATGAGGTCAAGGGCTG 608
Db 2351 GTACAAGGATGGCATGGAGCTGGGCCACTCCGGTGAAGCGCTTCTTGCAGGAGGATGTGGG 2410
Qy 609 CACAGC--AAGGCTGTGTGACAGAGTGGGCAAGCAGATGCTGGGGAGTACAGCTG 665
Db 2411 GACCGGCAACGGCTGTGGCAGCCACAGTCCACAGGAGGATGAAGCAGCTACTCTG 2470
Qy 666 CGAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCATCACAGAGCCCAAGGAGTGT 725
Db 2471 CCGGTGGGCGAGACTCTGTGGACTTCGGCTCCGGCTCTCTGAGCCCAAGGCGGTGT 2530
Qy 726 TGCCAAGGACAGTGTGTGCATAATGAGTCCGAGCTGAGGACGGGGCCAGTGCACACT 785
Db 2531 TGCCAAGGACAGCCCGCGTGCAGGAGGTGACGGCTGAGGTGGGGGCCAGGCCACGCT 2590
Qy 786 GAGCTGTGAGTGGCCCGAGGCCAGACAGAGGTGACGTGGTGTACAAGGATGGGAAGCT 845
Db 2591 GAGCTGTGAGTGGCCCGAGGCCAGACAGATGAGGTGACGTGTACAAGGAGGAGAGT 2650
Qy 846 GAGCTCCAGTTCGAAAGTGCAGATAGAGGCTGCGGGCTGCATGCGGCAGCTGGTGTGCA 905
Db 2651 GAGCTCCAGTTCGAAAGTGCAGCTGGAGGCTGTGGGCTGTATGCGGAGGCTGTGTGCA 2710
Qy 906 GCAGGAGGCCAGCAGATGCTGGGGAGTACACTGTGAGGCTGGGG 953
Db 2711 GCAGGTGGGCCAGGCAGACTCCGGATAGTACAGCTGTGAAGCCAGGTG 2758

RESULT 5

US-10-027-632-272054
; Sequence 272054, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129


```
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272054
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-272054

Query Match      18.5%; Score 185.6; DB 13; Length 578;
Best Local Similarity 87.5%; Pred. No. 3.1e-41;
Matches 203; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 153 CACAGAGCCCAAGATGATGTTTGCAGAGGAGCAGTCAGTGCATATGATGATGAGGTGCAGGCTGA 212
DB 347 CCCAGAGCCCAAGTGGTGGTTTGCAGAGGAGCAGCCACACAGAGGAGGTGCAGGCTGA 406

QY 213 GCGGGGGGCCAGTGCCTGCTGAGCTGAGTGGCCCGCCAGCCAGCCAGAGGAGGTGCAGGCTG 272
DB 407 GCGGGGGGCCAGTGCCTGAGTGGCCCGCCAGCCAGCCAGAGGAGGTGCAGGCTG 466

QY 273 GTACAAGATGGGAGAGCTGAGTCCAGCTCAAAAGTGGGCATGAGGTGCAAAAGGCTG 332
DB 467 GTACAAGATGGGAGAGCTGAGTCCAGCTCCAAAGTGGCGTGGAGGCGGTGGGCTG 526

QY 333 CACAGGAGGCTGCTGCTGCACAGCGGGGCAAGCAGATGCTGGGGAGTAC 384
DB 527 CACAGGAGGCTGCTGCTGCAGCAGCGGGGCGGCGGAGGCTGAGTAC 578

RESULT 6
US-10-027-632-272054
; Sequence 272054, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272054
```

```
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-272054

Query Match      18.5%; Score 185.6; DB 17; Length 578;
Best Local Similarity 87.5%; Pred. No. 3.1e-41;
Matches 203; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 153 CACAGAGCCCAAGATGATGTTTGCAGAGGAGCAGTCAGTGCATATGATGATGAGGTGCAGGCTGA 212
DB 347 CCCAGAGCCCAAGTGGTGGTTTGCAGAGGAGCAGCCACACAGAGGAGGTGCAGGCTGA 406

QY 213 GCGGGGGGCCAGTGCCTGCTGAGCTGAGTGGCCCGCCAGCCAGCCAGAGGAGGTGCAGGCTG 272
DB 407 GCGGGGGGCCAGTGCCTGAGTGGCCCGCCAGCCAGCCAGAGGAGGTGCAGGCTG 466

QY 273 GTACAAGATGGGAGAGCTGAGTCCAGCTCAAAAGTGGGCATGAGGTGCAAAAGGCTG 332
DB 467 GTACAAGATGGGAGAGCTGAGTCCAGCTCCAAAGTGGCGTGGAGGCGGTGGGCTG 526

QY 333 CACAGGAGGCTGCTGCTGCACAGCGGGGCAAGCAGATGCTGGGGAGTAC 384
DB 527 CACAGGAGGCTGCTGCTGCAGCAGCGGGGCGGCGGAGGCTGAGTAC 578

RESULT 7
US-09-726-643-22
; Sequence 22, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-643-22

Query Match      12.3%; Score 123.2; DB 9; Length 1645;
Best Local Similarity 49.5%; Pred. No. 5.6e-24;
Matches 376; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

QY 230 TGCTGAGCTGTGAGGTGGCCGCCAGCCAGCGGAGGTGACGTGGTACAAAGATGGGAAGA 289
DB 591 TGCTGACTTGTGAGCTCTCAAGGTGGACTTCCCGGCNACCTGTGTACAAAGATGGGCAGA 650

QY 290 AGCTGAGCTCCAGTCAAAAGTGGGCATGAGGTGCAAAAGGTGCACACGGAGGTGGTGC 349
DB 651 AGTGGAGGAGAGCGAGTTGCTGGTGGTGAAGATGGATGGGCGCAACACACCGTCTGTATCC 710

QY 350 TGCCACAGCGGGGCAAGAGCAGATGCTGGGGAGTACAGTGTGAGGCTGGGGCCAGAGAG 409
DB 711 TGCCTGAGGCCAAAGTCCAGGACAGTGGCGGAGTTGAGTGCAGGACAGAAAGGGGTCTCGG 770

QY 410 TCTCCTTCCACTGCATCATACAGAGCCCAAGGGGTGTTTTCGAAAGAGAGCAGTCAAGTGC 469
DB 771 CCTTCTTCGGGCTCACTGTCCAAAGATCCTCCGTGCACATCG---TGCACCCCGAGAAC 827

QY 470 ATAATGAGGTGCAGGCTGAGCGGGGACCACTGCCATGCTGAGTGTGAGGTGGCCAGC 529
DB 828 ATGTGTTTCGTGCATGCCATAACTTCCGAGTGTGTCTGCTGCTGCTGAGGTGGAGCCGAG 887

QY 530 CCCAGAGCGAGGTGACGTGGTGTACAAGGACGGGGAAGAGCTGAGTCCAGTCCAAAGTAC 589
```

Db 888 AGGACGCCCTGTGCGTTGGTACAAAGGACGGGCGAGGAGTGGAGGAGTGAATTCGTGG 947
Qy 590 GCATGGAGTCAAGGCTGCACACAAAGCTGGTGTAGTCAGCAGGTGGCAAGCAGATG 649
Db 948 TGCTGGAGAAATGAGGGGCCCATCGCCCGCTGGTGTGCGCCGCCACCCATCCCTCAGACG 1007
Qy 650 CTGGGAGTACAGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTTCAACTGCACATCACAG 709
Db 1008 GGGGCGAGTTTCAGTGCCTGCTGGAGATGATGTGCGCTACTTCACTGTCAACATCACAG 1067
Qy 710 AGCCCAAGGAGTGTTCAGGAGGAGTGGTGGTGCATAAATAGAGTGGGAGTGCAGAGCAG 769
Db 1068 A---CGTCTCTCGTGGATCGTGTATCCAGCGCAAGGTGTATGTGGCAGCGGTGCGCC 1124
Qy 770 GGGCAGTGCACACTGAGCTGAGGTGGTGGCCAGGCCAGACAGAGTGCAGTGGTACA 829
Db 1125 TGGAGCGTGTGGTGTGACCTGTGAGCTATGCGCGCCCTGGGCGAGAGTGGCTGGACCA 1184
Qy 830 AGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTCCGCATAGAGGCTCGCGGCTGCATGC 889
Db 1185 AGGATGGAGAGGAGTGTGGAGGCCCGCGCTGCTCTGCAGAAAGACACTGTCC 1244
Qy 890 GGCAGCTGGTGTGCAGCAGGAGGCCAGGCAGATGCTGGGAGTACACCTGTGAGGCTG 949
Db 1245 GCGGCTGGTGTGCGCGCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAAATTG 1304
Qy 950 GGGGCCAGCGCTCTCTTCCACCTGGATGTTTCAGAGCC 989
Db 1305 ACGATGAGTGGCGCTCTTCACTGTACCGTCAACAGAGTGC 1344

RESULT 8

US-10-042-141-22
; Sequence 22, Application US/10042141
; Publication No. US20020183503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/10/042,141
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-141-22

Query Match 12.3%; Score 123.2; DB 13; Length 1645;
Best Local Similarity 49.5%; Pred. No. 5.6e-24;
Matches 376; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

Qy 230 TGCTGAGCTGTGAGTGGCCCGCCAGAGCGGAGGTGACGTGTGTTACAAAGATCGGAAGA 289
Db 591 TGCTGACTGTGAGCTCTCAAGGGTGGACTTCCCGGCAACCTGGTACAAAGATGGGCAGA 650
Qy 290 AGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGGTCAAGAGGTGCACAGGAGGTGTGTC 349
Db 651 AGGTGGAGGAGCGAGTGTGCTGGTGAAGATGGATGGGCGCAAAACACCGTCTGTATCC 710
Qy 350 TGCCACAGCGGGCAAGACAGATGCTGGGAGTACAGCTGTGAGCTGGGGGCCAGAGAG 409
Db 711 TGCTGAGGCCAAAGTCCAGACAGTGGCGAGTTTGTAGTGCAGGACAGAGGGGTCTCGG 770
Qy 410 TCTCTTCCACTGCACATCAAGAGCCCAAGGGGGTGTGTTGCGAAGAGCAGTCAAGTGC 469

Db 771 CCTTCTTCGGCGTCACTGTCCAAGATCCTCCGCTGCACATCG---TGGACCCCGAGAAC 827
Qy 470 ATATAGAGTGCAGGCTGAGCGGGACCACTGCCATGCTGAGCTGTGAGGTGCGCCAGC 529
Db 828 ATGTGTTCTGTCATGCCATTAACCTTCGAGTGTGTCACTGCTGGCTGTGAGGTGACCGAG 887
Qy 530 CCACAGACGAGGTCACCTGGTACAAGACGGAAGAGCTGAGCTCCAGCTCAAAAGTAC 589
Db 888 AGGACGCCCTGTGCTGGTTGGTACAAGACGGGAGGAGGTGGAGGAGTGAATTCGTGG 947
Qy 590 GCATGGAGTCAAGGCTGCACACAAAGCTGGTGTAGTGCAGCAGGTGGGCAAGCAGATG 649
Db 948 TGCTGGAGAAATGAGGGGCCCATCGCGCGCTGGTGTGCGCCGCCACCCATCCCTCAGACG 1007
Qy 650 CTGGGAGTACAGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTTCAACTGCACATCACAG 709
Db 1008 GGGGCGAGTTTCAGTGCCTGCTGGAGATGAGTGTGCTTCACTGTCAACATCACAG 1067
Qy 710 AGCCCAAGGAGTGTGTTTGGCCAAAGGAGCAGTGTGGTGCATAATAGGTGCGGACTCAGGCAG 769
Db 1068 A---CGTCTCTCTGATCGTGTATCCAGCGCAAGGTGTATGTGGCAGCGGTGCGCC 1124
Qy 770 GGGCAGTGCACACTGAGCTGTGAGTGGCCAGGCCAGACAGAGGTGACGTGGTACA 829
Db 1125 TGGAGCGTGTGGTGTGACCTGTGAGCTATGCGCGCGCTGGGCGAGGTTGCGTGGACCA 1184
Qy 830 AGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTCGGGCTGCATGC 889
Db 1185 AGGATGGAGAGGAGTGTGGAGAGCCCGCGCTGCTCTGCAGAAAGACACTGTCC 1244
Qy 890 GGCAGCTGGTGTGCAGCAGGCGCAGGCAGATGCTGGGAGTACACCTGTGAGGCTG 949
Db 1245 GCGGCTGGTGTGCTGCGCGCTGCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAAATTG 1304
Qy 950 GGGGCCAGCGGCTCTCTTCCACCTGGATGTTTCAGAGCC 989
Db 1305 ACGATGAGTGGCGCTCTTCACTGTCAACCGTCAACAGAGTGC 1344

RESULT 9

US-10-919-272-22
; Sequence 22, Application US/10919272
; Publication No. US20050010042A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/10/919,272
; CURRENT FILING DATE: 2004-08-17
; PRIOR APPLICATION NUMBER: US/09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-919-272-22

Query Match 12.3%; Score 123.2; DB 19; Length 1645;
Best Local Similarity 49.5%; Pred. No. 5.6e-24;
Matches 376; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

Qy 230 TGCTGAGCTGTGAGTGGCCCGCCAGAGCGGAGGTGACGTGTGTACAAAGATCGGAAGA 289
Db 591 TGCTGACTGTGAGCTCTCAAGGGTGGACTTCCCGGCAACCTGGTACAAAGATGGGCAGA 650
Qy 290 AGCTGAGCTCCAGCTCAAAAGTGGGCATGGAGGTCAAGGGTGCACACGAGGCTGTGTC 349

Db 651 AGGTGGAGGAGCCAGTTGCTGGTGAAGATGGATGGGCGCAACACCGTCTGTATCC 710
Qy 350 TGCCACAGCGGGGCAAGACAGATGCTGGGGAGTA CAGCTGTGAGCTGGGGCCAGAGAG 409
Db 711 TGCTGAGGCCAAAGTCCAGGACAGTGGCGAGTTTGTAGTGCAGGACAGAAAGGGTCTCGG 770
Qy 410 TCTCTCTCCACTGCACATCAGAGCCCAAGGGGTGTTTCCGAAGGACAGTCAAGTGC 469
Db 771 CTTCTTTCGGCGTCACTGTCCAAAGATCCTCCCGTGCACATCG---TGGACCCCGAGAAC 827
Qy 470 ATAATGAGGTGAGGCTGAGGCGGGGACCACTGCCATGCTGAGCTGTGAGGTGGGCCACAGC 529
Db 828 ATGTGTTGCTGATGCCATACTTCGGAGTGTGTATGCTGGCTGTGAGGTGGACCGAG 887
Qy 530 CCCAGACGGAGGTGAGTGGTGAACAAGGACGGGAAGAGTCCAGCTCAAAAGTAC 589
Db 888 AGGACGCCCTGTGCGTGGTGAAGGACGGGACAGGAGTGGAGAGTGAAGTCTCGTGG 947
Qy 590 GCATGGAGTCAAGGCTGCACACGAAGGCTGGTAGTCAGCAGAGTGGGCAAGCAGATG 649
Db 948 TGCTGGAGATGAGGGGCCCCATCCCGCTGGTCTGCGCGCCACCCATCCCTCAGACG 1007
Qy 650 CTGGGAGTACAGCTGCGAGGCTGGGGCCAGAGAGTCTCCTTTCAACTGCACATCACAG 709
Db 1008 GGGGCGAGTTTCACTGCGCTGAGAGTGTGAGTGTGCTTCACTGTCAACATCACAG 1067
Qy 710 AGCCCAAGCAGTGTTCACGAAGGAGCAGTTGGTGCATAAATGAGTGGGAGTCAAGGACG 769
Db 1068 A---CGTCTCTCGTGGATCGTGTATCCAGCGCAAGGTGTATGTGGCAGCGTGCGCC 1124
Qy 770 GGGCAGTGCACACTGAGCTGTGAGTGGGCCAGGCCACAGACAGAGTGTGAGTGTGACA 829
Db 1125 TGGACGCTGTGTGCTGACCTGTGAGCTATGCGCGCCCTGGGACAGAGTGGCGTGACCA 1184
Qy 830 AGGATGGGAAGAGTGTGAGCTCCAGTTCGAAGTCCGATAGAGCTGCGGGCTGCATGC 889
Db 1185 AGGATGGAGAGAGTGTGTGAGAGCCCGCGCTGCTCTGCAGAAAGACATGTGCC 1244
Qy 890 GGCAGCTGGTGTGACAGCAGGAGCCAGGACAGATGCTGGGAGTACACCTGTGAGGCTG 949
Db 1245 GCGCGCTGGTGTGCGCGCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAATTTG 1304
Qy 950 GGGGCCAGCGGCTCTCTTCCACCTGGAGTGTTCAGAGCC 989
Db 1305 ACGATGAGTCGGCCCTCTTCACTGTCAACGTCACAGAGTGC 1344

RESULT 10

US-10-264-049-850
; Sequence 850, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birste et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 850
; LENGTH: 2170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-049-850

Query Match 12.3%; Score 123.2; DB 17; Length 2170;
Best Local Similarity 49.5%; Pred. No. 5,8e-24;
Matches 376; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

Qy 230 TGCTGAGCTGTGAGTGGCCCCCAGACGGAGGTGACGTGTGTACAAAGNTGGGAAGA 289
Db 1082 TGCTGACTTGTGAGCTCTCAAGGGTGGACTTTCCCGGCAACCTGGTGTACAAAGGTGGGCGAGA 1141
Qy 290 AGCTGAGCTCCAGCTCAAAAGTGGGCAATGGAGGTCAAAGGGTGCACACGGAGGCTGGTGC 349
Db 1142 AGGTGGAGGAGCGAGTTGCTGTGGTGAAGATGGATGGGCGCAAAACACCGTCTGTATCC 1201
Qy 350 TGCCACAGCGGGGCAAAAGCAGATGCTGGGGAGTACAGCTGTGAGCTGGGGGCCAGAGAG 409
Db 1202 TGCTGAGGCCAAAGTCCAGGACAGTGGCGAGTTTGTAGTGCAGGACAGAAAGGGTCTCGG 1261
Qy 410 TCTCTTCCACTGCACATCAGAGCCCCAAGGGGTGTTTTCGGAAGGAGCAGTCAAGTGC 469
Db 1262 CTTCTTTCGGCGTCACTGTCTCAAGATCCTCCCGTGCACATCG---TGGACCCCGAGAAC 1318
Qy 470 ATAATGAGGTGAGGCTGAGGCGGGGACCACTGCCATGCTGAGCTGTGAGGTGGGCCACAGC 529
Db 1319 ATGTGTTGCTGCATGCCATAACTTCCGAGTGTGTATGCTGGCTGTGAGGTGGACCGAG 1378
Qy 530 CCCAGACGGAGTGTGAGTGTGACGAAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTAC 589
Db 1379 AGGACGCCCTGTGCTTTGGTACAAGGACGGGAGGAGTGGAGGAGTGAATTCGTTGG 1438
Qy 590 GCATGGAGGTCAAGGCTGCACACGAAGGCTGGTAGTCAGCAGAGTGGGCAAGCAGATG 649
Db 1439 TGCTGGAGAAATGAGGGGCCCCATCGCGCTGGTGTGCTGCGCGCCACCCAGCCCTCAGACG 1498
Qy 650 CTGGGAGTACAGCTGCGAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAG 709
Db 1499 GGGGCGAGTTTCACTGCTGCTGTGAGATGATGTGCTGCTTCACTGTCAACATCACAG 1558
Qy 710 AGCCCAAGCAGTGTTCGCCAAGGAGCAGTTGGTGTGCATAAATGAGTGGGAGTCCAGGCGAG 769
Db 1559 A---CGTCTCTCGTGGATCGTGTATCCAGCGCAAGGTGTATGTGGCAGCGTGCGCC 1615
Qy 770 GGGCAGTGCACACTGAGCTGTGAGTGGGCCAGAGAGTCTCTTTCAACTGCACATCACAG 829
Db 1616 TGGACGCTGTGCTGCTGACCTGTGAGCTATCCCGCCCTGGGCGAGGTTGGCTGGACCA 1675
Qy 830 AGGATGGGAAGAGTGTGAGCTCCAGTTCGAAGTTCGCATAGAGGCTCGGGCTGCATGC 889
Db 1676 AGGATGGAGAGAGTGTGTGAGAGCCCGCGCTGCTCTCTGCAGAAAGACACTGTGCC 1735
Qy 890 GGCAGCTGGTGTGACAGCAGGACCGCCAGGAGATGCTGGGGAGTACACCTGTGTAGGCTG 949
Db 1736 GCGCGCTGGTGTGCGCGCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAATTTG 1795
Qy 950 GGGGCCAGCGGCTCTCTTCCACCTGGAGTGTTCAGAGCC 989
Db 1796 ACGATGAGTCGGCCCTCTTCACTGTCAACGTCACAGAGTGC 1835

RESULT 11

US-10-093-463-73
; Sequence 73, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennnda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li

APPLICANT: Zerhusen, Bryan
APPLICANT: Tchernev, Velizar
APPLICANT: Gangolli, Esha
APPLICANT: Vernet, Corine
APPLICANT: Pena, Carol
APPLICANT: Burgess, Catherine
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly
APPLICANT: Gorman, Linda
APPLICANT: Spaderna, Steven
APPLICANT: Voss, Edward
APPLICANT: Malyankar, Uriel
APPLICANT: Anderson, David
APPLICANT: Patturajan, Meera
APPLICANT: Miller, Charles
APPLICANT: Taupier, Raymond J. Jr.
TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypepti
TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.
FILE REFERENCE: 21402-290A (Cura 590AT)
CURRENT APPLICATION NUMBER: US/10/093,463
PRIORITY FILING DATE: 2002-06-24
PRIORITY FILING DATE: 2001-04-14
PRIORITY FILING DATE: 2001-12-03
PRIORITY FILING DATE: 2001-03-08
PRIORITY FILING DATE: 2001-03-08
PRIORITY FILING DATE: 2001-03-08
PRIORITY FILING DATE: 2001-09-27
PRIORITY FILING DATE: 2001-07-10
PRIORITY FILING DATE: 2001-07-10
PRIORITY FILING DATE: 2001-03-30
PRIORITY FILING DATE: 2001-05-31
PRIORITY FILING DATE: 2001-04-30
PRIORITY FILING DATE: 2001-06-18
PRIORITY FILING DATE: 2001-07-31
PRIORITY FILING DATE: 2001-04-04
PRIORITY FILING DATE: 2001-03-08
PRIORITY FILING DATE: 2001-03-09
PRIORITY FILING DATE: 2001-10-18
PRIORITY FILING DATE: 2001-03-12
PRIORITY FILING DATE: 2001-05-03
PRIORITY FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 370
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 73
LENGTH: 14061
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (15)..(14039)
US-10-093-463-73

Query Match 9.7%; Score 97.4; DB 17; Length 14061;
Best Local Similarity 46.5%; Pred. No. 9.3e-17;
Matches 437; Conservative 0; Mismatches 481; Indels 21; Gaps 3;
QY 48 CAAGGCTGCAGACGGAGGCTGGTGTGCACAGCAGCGCAAGACGGATGCCGGGGACTA 107

Db 53 CAAGGGCGGGCCACACGCTGGTACTGAAGAGCATCCAGCGGGCTGATCGGGCATAGT 112
QY 108 CAGCTGCGAGGCCAGGGGCCAGAGGGTCTCTCCGGCTGCACATCACAGAGCCCAAGAT 167
Db 113 ACGCGCTCTCTCCTGAAGGTGTCGACCTCTGCGCGCTGGAGGTCCAGGTGAAGCCGGT 172
QY 168 GATGTTTGCAAAGGAGCAGTCAGTGCATATAGAGGTGCAGGCTCAGGCGGGGCCAGTGC 227
Db 173 GGTGTTCTGAAGGGCTG-----GATGACTGTCCGACAGAGGAGCGGGCACCT 223
QY 228 CATGCTGAGCTGTGAGGTGGCCAGGCCCAAGCGAGGTGACGTGGTACAGAGTGGAA 287
Db 224 GGCCCTGCAGTGTGAAGTCTCTGACCCCGAGGCCATGTGGTGTGGGCCCAAGATGGCGT 283
QY 288 GAAGCTGAGTCCAGCTCAAAAGTGGGCATGGAGGTCAAAAGGTGCACACGAGAGCTGGT 347
Db 284 GAGCTGGGCCCCAGTGAAGATATGACTTCTCTCACACGCGGGGCACGCGGGGGCTCGT 343
QY 348 GCTCCACAGCGGGCCAAAGCAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGGCCAGAG 407
Db 344 GGTGCATGACGTGAGCCCTGAAGAGCGCGGCTGTACACCTGCCACGCTGGGCTCCGAGGA 403
QY 408 AGTCTCTTCACTGACATCAACAGAGCCCAAGGGGGTGTTCGCAAGAGAGTACGT 467
Db 404 GACCGGGCGGGTCCGCGTGCACGATCTGCAGTGGGCATCAACCAAGAGGCTG---AA 460
QY 468 GCATAATGAGTGCAGGCTGAGCGGGGACCACTGCCATGCTGAGCTGTGAGGTGGGCCA 527
Db 461 GACATGAGGTGCTGGAAGGGGAAAGCTGACAGCTTTGAGTGGCTCTCTGCCACGAGAG 520
QY 528 GCCCAGACGAGGTGACGTGGTACAAGAGACGGGAAGAGCTGAGCTCCAGCTCAAAAGT 587
Db 521 TGCACAGCACCGGCCCATGTGGACAGTGGTGGGAAGACAGTGGGAGCTCCAGCCGCTT 580
QY 588 ACGCATGGAGGTCAAGGGTGCACACGAAGGGTGTGTAGTGCAGCAGGTGGGCAAGCAGA 647
Db 581 CCAGGCCACACGTCAGGGCCGAAATACATCTGTGTCCGGGAGGCTGCACCAAGTGA 640
QY 648 TGCTGGGGAGTACAGCTGCGAGGCTGGGGGCCAGAGAGTCTCTTTCAACTGCACATCAC 707
Db 641 TGCGGGGAGGTGGTCTTCTGTGCGGGGCTTCACTCAAGGCTCTCTCATTTGTAG 700
QY 708 AGAGCCCAAGGCAGTGTTCGCAAGGAGCAGTGTGGTGCATATAGAGTGGGACTGAGGC 767
Db 701 AGAGAGCGCGCGCCCATCATCAAGCCCTGGAAGACCAGTGGGTGGCGCCAGGGAGGA 760
QY 768 AGGGCCAGTGCACACTGAGCTGTGAGGTGGCCAGGCCACAGCAGAGGTGACGTGGTA 827
Db 761 CGTGA-----GCTGCGCTGTGAGCTGTCAAGGACGCGGGGAAACGCCCTGCACTGGCT 811
QY 828 CAAGGATGGGAAGAAGCTGAGCTCCAGTTCGAAAGTCCGCATAGAGGTGCGGGCTGCAT 887
Db 812 GAAGGACAGGAAGGCCATCCGCAAGAGCCAGAGATATGATGTGTCTGCGAGGGCAGAT 871
QY 888 GCGCAGCTGTGTGTGCAGCAGGCGAGGCCAGGAGATGCTGGGAGGTACACTGTGAGGC 947
Db 872 GGCCATGCTGTCTCATCCGCGGGGCTCGCTCAAGGACGCGGGGAGTACACGTGTGAGGT 931
QY 948 TGGGGGCCAGGGCTCTCTTCCACCTGGATGTTTCAGA 986
Db 932 GGAGGCTTCCAAGAGCACAGCCAGCCCTCCATGTGGAAGA 970

RESULT 12
US-10-093-463-71
; Sequence 71, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir


```
RESULT 13
US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1946

Query Match          9.16%; Score 95.8; DB 17; Length 2534;
Best Local Similarity 46.4%; Pred. No. 2.1e-16;
Matches 436; Conservative 0; Mismatches 482; Indels 21; Gaps 3;

QY 48 CAAGGCTGCAGACGAGAGGCTGTGTGCAGAGGACGAGCAGGATGCCGGGACTA 107
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 129 CAAGGGCCGGCGCCACACGCTGTACTGAAGAGCATCCAGCGGGCTGATCGGGCATAGT 188
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 108 CAGCTGCGAGGCCAGGGCCAGAGGCTCTCTTCGCTGCACATCAGAGCCCAAGAT 167
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 189 ACGGCCCTCTCCCTGAAGGTTGTCACCTCTGCCCGCTGGAGTCCAGGTGAAGCCGGT 248
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 168 GATGTTTCAAGGAGCAGTCAAGTATGAGTGCAGGCTGAGGCGGGGCGCAGTGC 227
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 249 GGTGTTCTGAAGCGCTG-----GATGACCTGTCGACAGAGGAGCGGCACCCCT 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 228 CATGCTAGCTGTGAGGTGGCCAGGCCAGCGAGGTGATGCTGTGTAAGAGTGGAA 287
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 GGCCTGCACTGTGAAGTCTCTGACCCCGAGGCCCATGTGTGTGGCCAAAGATGGCGT 359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 288 GAAGCTAGCTCCAGCTCAAAAGTGGGATGGAGTCAAGGTCGACACGAGGCTGGT 347
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 360 CGAGCTGGGCCCCAGTGAAGATGATCTCTGACACGGCGGGCGACGCGGGGCTCGT 419
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 348 GCTGCCACAGGGCGGCAAGCAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGCCAGAG 407
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 420 GGTGATGACGTGAGCCCTGAAGCGCGCTGTACACCTGCCACGTGGGCTCGAGGA 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 408 AGTCTCTTCCACCTGCATCAGAGCCCAAGGGGTGTTTGGGAAGAGAGTCACTAGT 467
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 480 GACCGGGCCGGGTCGCGCTGCAGATCTGCACGTGGGCGCATCCCAAGAGGCTG---AA 536
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 468 GCATAATGAGGTGAGGCTGAGGCGGGGACCACTGCCATGCTGAGCTGTGAGGTGGCCCA 527
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 537 GACATGAGGTGCTGGAAGGGGAAAGCTGCAGCTTTTGAAGTGGCTCTGTCTCCACAGAG 596
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 528 GCGCCACAGGAGGTGACGTGTGTAAGAGCGGGAAGAGTGTGAGCTCCAGCTCAAAAGT 587
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 597 TGCCAGGACCCGGCCATGTGACAGTCGTTGGGAGACAGTGGGACGCTCCAGCCGCTT 656
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 588 ACGATGAGGTCAAGGCTGCACAGAGGCTGTGTAGTGACAGAGTGGGCAAGAGAG 647
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 657 CCAGGCCACAGCTCAGGGCCGAAAATACATCTGTGTGTCGGGAGGCTGCACCAAGTGA 716
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 648 TGCTGGGAGTACAGTSCGAGGCTGGGGGCCAGAGTCTCTTCACTGCACATCAC 707
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 717 TGCCGGGAGGTGCTTCTCTGTGCGGGGCTCACCTCAAGGCTCACTCATTTGTAG 776
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 708 AGAGCCCAAGGAGTGTGTCAGAGGAGCAGTGTGTGATATAGGTGCGGACTGAGGC 767
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 777 AGAGGGCCGCCACCATCATCAAGCCCTTGAAGACCAGTGGGTGGCGCCAGGGAGGA 836
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 768 AGGGCCAGTCCACACTGAGCTGTGAGTGGGCCAGGCCCCAGACAGAGGTGACGTGTA 827
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
US-10-476-397-15
; Sequence 15, Application US/10476397
; Publication No. US20040115687A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: LEE, Ernestine A.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: DING, Li
; APPLICANT: JACKSON, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: LEE, Sally
; APPLICANT: WARREN, Bridget A.
; APPLICANT: XU, Yuming
; APPLICANT: TRAN, Uyen K.
; APPLICANT: LAL, Preeti G.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: HAPALIA, April J.A.
; APPLICANT: YAO, Monique G.
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: CHINN, Anna M.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: FORSYTHE, Ian J.
; TITLE OF INVENTION: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS
; FILE REFERENCE: PP-0968 USN
; CURRENT APPLICATION NUMBER: US/10/476,397
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: PCT/US02/13874
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/288,290
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/292,468
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/298,616
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/301,672
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/345,008
; PRIOR FILING DATE: 2002-01-04
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 7564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7326129CB1
US-10-476-397-15
```

```
Query Match          9.6%; Score 95.8; DB 18; Length 7564;
Best Local Similarity 46.4%; Pred. No. 2.4e-16;
Matches 436; Conservative 0; Mismatches 482; Indels 21; Gaps 3;

QY 48 CAAAGGCTGCAGACGAGGCTGTGGTGCAGCAGGACGAGCAGGATCCCGGGGACTA 107
Db 1938 CAAAGGCGCGCCACACAGCTGTACTGAAGAGCATCCAGCGGGTGATCGGGCATAGT 1997
QY 108 CAGCTGCAGAGCGCAGGGGCGAGAGGCTCTCTTCGCTGTGCATCAACAGAGCCCAAGAT 167
Db 1998 ACGCCCTCTCTCCCTGAAGGTTGCACCTTCGCCGCTGGAGTCCGAGTGAAGCCGT 2057
QY 168 GATGTTTGCAGAGGAGCTAGTGCATATAGCTGAGCTGAGGCGGGGCGCAGTGC 227
Db 2058 GGTGTTTCTGAAGGCGCTG-----GATGACCTGTCCGAGAGGAGCGCGCACCCCT 2108
QY 228 CATGCTGAGCTGTGAGGTGGCCAGCCAGACGAGGTGACGTGTGTACAGAGTGGGAA 287
Db 2109 GGCCCTGCAGTGTGAAGTCTCTGACCCCGAGGCCCATGTGTGTGCGCAAGATGGCGT 2168
QY 288 GAAGCTGAGTTCACGCTCAAAAGTGGGATGGGATGAAAGGTGCACACGAGGCTGGT 347
Db 2169 GCAGCTGGGGCCCCAGTGCACAGTATGACTTCTGTGCACACGCGGGGCGACGCGGGGGCTGGT 2228
QY 348 GCTGCCACAGCGCGGCAAGCAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGGCGAG 407
Db 2229 GGTGCATGACGTGAGCCCTGGAAGACGCGCGGCTGTACACCTGCCACATGGGGTCCGAGGA 2288
QY 408 AGTCTCTTCCACCTGCACATCACAGAGCCCAAGGGGTGTTTGGCAAGGAGCAGTCAGT 467
Db 2289 GACCGCGGCCGGGTCCGGGTGCACGATCTGCACGTGGGCATCACCAAGAGCTG---AA 2345
QY 468 GCATATGAGGTGACGTGAGGTGAGCGGGGACCACTGCCATGTGTGAGCTGTGAGGTGGCCCA 527
Db 2346 GACAAATGAGGTGCTGGAAGGGGAAAGCTGCAGCTTTGAGTGCCTCTGTCCACGAGAG 2405
QY 528 GCCCAGACGAGGTGAGTGTGTACAGGACGGGAGAGCTGAGCTCCAGTCAAAGT 587
Db 2406 TGCCAGACGACCGGGCCATGTGACAGTGTGGTGGGAGAGCAGTGGGCGAGCTCCAGCGGTT 2465
QY 588 ACGCATGAGGTCAAGGGCTGCACACGAAAGGCTGTGAGTGCAGCAGGTGGGCAAGCAGA 647
Db 2466 CCAGGCCACAGCTCAGGGCCGGAATAATCATCTGTGTGTCCGGGAGGCTGCACCAAGTGA 2525
QY 648 TGCTGGGAGTACAGCTCGAGGCTGGGGGCGCAGAGAGTCTCTTTTCAACTGCACATCAC 707
Db 2526 TGCCGGGAGGTGGTCTTCTGTGCGGGGCTCACCTCCAAGGCTCACTCATTTGTGAG 2585
QY 708 AGAGCCCAAGCAGTGTGTCAGGAGCAGTGTGTGATATATGAGTGCAGGACTGAGGC 767
Db 2586 AGAGAGCGCGCGCCCATCATCAGCCCTTGAAGACCAAGTGGGTGGCGCCAGGGGAGGA 2645
QY 768 AGGGGCCAGTGCACACTGAGCTGTGAGTGGCGCCAGGCCACAGACAGAGGTGACGTGTGA 827
Db 2646 CGTGA-----GCTGCGCTGTGAGCTGTACAGGGGGGAAAGCCCGGTGACATGGCT 2696
QY 828 CAAGATGCGGAAGAGCTGAGTCCAGTTCGAAAGTGCGCATAGAGGCTGCGGGCTGCAT 887
Db 2697 GAAGGACAGGAAGGCCATCCGCAAGAGCCAGAAGTATGATGTGGTCTGCGAGGGCACCAT 2756
QY 888 CGGCGAGCTGTGTGTGACGAGGAGGCGAGGCAGATGTGTGGGAGTACACCTGTGAGGC 947
Db 2757 GGCCATGTGTGTATCCCGGGGCTCTGCTCAAGGACGCGGGCGAGTACAGCTGTGAGGT 2816
QY 948 TGGGGGCCAGCGGCTCTCTTCCACTGATGTTTTCAGA 986
Db 2817 GGAGCTTCCAGAGACACAGCCAGCTTCATGTGGAAGA 2855
```

RESULT 15

US-09-726-643-44

; Sequence 44, Application US/09726643

; Patent No. US20020028449A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 26 Human secreted proteins

; FILE REFERENCE: P2040P1

; CURRENT APPLICATION NUMBER: US/09/726,643

; CURRENT FILING DATE: 2000-12-01

; PRIOR APPLICATION NUMBER: PCT/US00/15187

; PRIOR FILING DATE: 2000-06-02

; PRIOR APPLICATION NUMBER: 60/137,725

; PRIOR FILING DATE: 1999-06-07

; NUMBER OF SEQ ID NOS: 190

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 44

; LENGTH: 2254

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-726-643-44

```
Query Match          8.8%; Score 88.4; DB 9; Length 2254;
Best Local Similarity 50.4%; Pred. No. 2.3e-14;
Matches 269; Conservative 0; Mismatches 261; Indels 4; Gaps 2;

QY 180 GGAGCAGTCAGTGCATAATGAGGTGCAGGCTGAGCGGGGGCCAGTGCATGCTGAGCTG 239
Db 1309 GGACCCCGAGAACATGTGTTGTCATGCCATAACTTCCGAGTGTGATGCTGCGCTG 1368
QY 240 TGAGTGTGCCCGAGGCCCGAGCGAGGTGACGTGTGTTACAAAGGATGGAAAGAACTGAGCTC 299
Db 1369 TGAGTGTGACCGAGAGGACGCCCTGTGCTGTGTTACAAAGGACGGGAGGAGGTGGAGGA 1428
QY 300 CAGCTCAAAAGTGGCATGGAGGTCAAAGGGTGACACGGAGGCTGGTGTGCCACAGGC 359
Db 1429 GAGTGACTTCGTGTGTGTGAGAGTGAAGGGGGCCCATCGCGGCTGGTGTGCTCGGCCAC 1488
QY 360 GGGCAAAACAGATGCTGGGGAGTACAGCTGTGAGGCTGGGGGCCAGAGAGTCTCTTTCCA 419
Db 1489 CCAGCCCTCAGACGGGGCGAGTTTCAGTGGTGTGAGGATGAGTGTGCTACTTTCAC 1548
QY 420 CTTGCATCATCAGAGCCCAAGGGGGTGTGTCGAAAGGAGCAGTCAAGTGCATAAATGAGGT 479
Db 1549 TGTCAACCATCACAGACGTCCTCTGTGATCGTGTAT---CCAGCGGCAAGGTGTATGT 1605
QY 480 GCAGGCTGAGCGGGGACCACTGCCATGCTGAGCTGTGAGTGGGCCAGCCCGCCAGACGGA 539
Db 1606 GGCAGCCGTGCGCCTGGAGCGGTGTGCTGACCTGTGAGCTATCGCGGCTCGGGCAGA 1665
QY 540 GGTGACGTGTGTACAGGACGGGAAGCTGAGCTCCAGCTCAAAAGTACGCATGAGGT 599
Db 1666 GGTGCGCTGGACCAAGGATGGAGAGGAGTG-GTGGAGACCCCGCGCTGCTCTGCAGNA 1724
QY 600 CAAAGGCTGCACAGAGGCTGGTAGTGCAGCAGGTGGGCAAGCAGATGTGTGGGAGTGA 659
Db 1725 GGAAGACACTGTCCGCGGCTGGTGTGCTGCCGTGTCTCCAGCTCGAGGACTCCGGCGAGTA 1784
QY 660 CAGTGCAGGCTGGGGGCCAGAGTCTCTTTCACTGCACATCAGAGGCC 713
Db 1785 CTTGTGTGAAATTGACGATGAGTGGGCTCTTCTACTGTCTACCTTCACAGAGTC 1838
```

Search completed: March 21, 2005, 16:07:33

Job time : 633.052 secs

THIRTY FOUR EIGHT (0080)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:02:07 ; Search time 175.678 Seconds
(without alignments)
9323.380 Million cell updates/sec

Title: US-10-077-130-4_COPY_10500_11500
Perfect score: 1001
Sequence: 1 gcccctgtgagtgaggaa.....ttcatagaagatgtgaaaaa 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	67.6	6.8	7218	1	US-08-232-463-14
2	46.4	4.6	767677	4	US-09-949-016-12147
3	46.4	4.6	767677	4	US-09-949-016-17361
C 4	42.4	4.2	1926	3	US-09-249-585A-4
C 5	42.4	4.2	1931	2	US-09-130-114-2
C 6	42.2	4.2	4276	4	US-09-949-016-4900
C 7	42.2	4.2	5065	4	US-09-949-016-744
C 8	42.2	4.2	50453	4	US-09-949-016-16642
C 9	42.2	4.2	51242	4	US-09-949-016-12486
C 10	41.6	4.2	767677	4	US-09-949-016-12147
C 11	41.6	4.2	767677	4	US-09-949-016-17361
12	41.2	4.1	289	3	US-09-007-005-17
13	41.2	4.1	289	3	US-09-244-796-17
14	40.8	4.1	601	4	US-09-949-016-61224
C 15	40.6	4.1	1166	3	US-09-072-596-323
C 16	40.6	4.1	1166	4	US-09-072-967-328
17	39.6	4.0	63930	4	US-09-949-016-12270
18	39.6	4.0	109690	4	US-09-949-016-13525
19	39.4	3.9	601	4	US-09-949-016-168436
20	39.4	3.9	43414	4	US-09-949-016-12839
21	39.4	3.9	43415	4	US-09-949-016-16491
C 22	39.2	3.9	965	1	US-08-736-361A-3
C 23	39.2	3.9	1576	1	US-08-736-361A-1
C 24	38.8	3.9	1975	4	US-09-308-345A-6
25	38.8	3.9	2295	4	US-09-949-016-5079
C 26	38.8	3.9	2456	3	US-08-813-150-5
C 27	38.8	3.9	2456	4	US-09-546-553-5

Query Match 6.8%; Score 67.6; DB 1; Length 7218;

Sequence 93, Appl
Sequence 93, Appl
Sequence 177691,
Sequence 358, App
Sequence 349, App
Sequence 16821, A
Sequence 35, Appl
Sequence 31, Appl
Sequence 35, Appl
Sequence 6866, Ap
Sequence 580, App
Sequence 2, Appl
Sequence 3, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 14, Appl
Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-F1s
US-08-232-463-14

Best Local Similarity 2.6%; Pred. No. 3.1e-07;
Matches 10; Conservative 237; Mismatches 141; Indels 0; Gaps 0;
QY 436 AGGGCTTACCCATCAAGTTACAGAGGGTCTGAGGAACGAGAGGCCACAGAGGGCA 495
Db 1455 AGAGATAGAAGATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1396
QY 496 ACAGCCGTCTCGGTGTGAGTGTGAGCAAGATGGCCCGCTGGAGTGTGGAAGGGCAT 555
Db 1395 RRR 1336
QY 556 GAGACCTCAGATGGAGACAGACAGCCCTGAGGAGAGCGGGCCAGGTGTGAGCTG 615
Db 1335 RRR 1276
QY 616 CAGATCCGCGCTCGTGGCAGAGACCTGGGAGTACCTGTGATGCGGGAAGAG 675
Db 1275 RRR 1216
QY 676 AGGACTCAGCATCTCACCGTCAGGCGCATGCTTCCAAAGTTTCATAGAGGCTGAGG 735
Db 1215 RRR 1156
QY 736 AATGAAGGCCACAGAGGGGACAGCGCCACGCTGTGTGTGAGCTGAGCAAGCGGCA 795
Db 1155 RRR 1096
QY 796 CCGTGGAGTGGAGGAAGGGGCATGAGA 823
Db 1095 RRR 1068

RESULT 2
US-09-949-016-12147
; Sequence 12147, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12147
; LENGTH: 767677
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(767677)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12147

Query Match 4.6%; Score 46.4; DB 4; Length 767677;
Best Local Similarity 45.0%; Pred. No. 0.27;
Matches 214; Conservative 0; Mismatches 261; Indels 1; Gaps 1;
QY 468 GAGGAACGAGAGGCCACAGAGGGGCAACAGCGCTGCTGCGGTGTGAGCTGACAGAT 527
Db 532241 GAAGAGGAGGAGAGAGGGGAGAGAGGAGAGAGAGGAGAGGAGAGGAGAGGA 532300
QY 528 GGCCCCCGTGGAGTGGTGAAGGGGCATGAGACCTCAGAGATGAGAGACAGACACCT 587
Db 532301 GAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 532360

QY 588 GAGCAGAGCGGGCCAGGTGTGAGCTGCAGATCCGGCGCCTCGTGGCAGAGGACGCTGG 647
Db 532361 GAGGAGGAGGAGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 532420
QY 648 GGAGTACCTGTGCATGTGCGGGAAGGAGAGACCTCAGCCATGTCTACCGTTCAGGCGCAT 707
Db 532421 GGAGGAGATGAGGAGGAGGAGTGTGAG-GAAGAGAGAGAGGAGGAGGAGGAGGAGGA 532479
QY 708 GCCTTCCAAAGTTTCATGAGAGGTCTGAGGAATGAAGAGGCCACAGAAAGGGGACAGCCAC 767
Db 532480 GGAGGAGGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532539
QY 768 GCTGTGTGTGAGCTGAGCAAGCGCGCACCGTGTGAGTGGAGGAAGGGGCATGAGACCT 827
Db 532540 GGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 532599
QY 828 CAGAGATGGGACAGACAGACCTGAGGAGGAGCGGTCCAGGTGTGAGCTGAGATCCG 887
Db 532600 GGAGGAGGAGGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532659
QY 888 TGGCCTCGCTGTGTGTGATCCCGGGAGTACTGTGTGTGTGCGGCGCAGGAGGAGGA 943
Db 532660 GGAGGAGTGTGAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTGTGAAAAGGAGGA 532715

RESULT 3
US-09-949-016-17361
; Sequence 17361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17361
; LENGTH: 767677
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(767677)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17361

Query Match 4.6%; Score 46.4; DB 4; Length 767677;
Best Local Similarity 45.0%; Pred. No. 0.27;
Matches 214; Conservative 0; Mismatches 261; Indels 1; Gaps 1;
QY 468 GAGGAACGAGAGGCCACAGAGGGGCAACAGCGCTGCTGCGGTGTGAGCTGACAGAT 527
Db 532241 GAAGAGGAGGAGAGAGGGGAGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 532300
QY 528 GGCCCCCGTGGAGTGGTGAAGGGGCATGAGACCTCAGAGATGAGAGACAGACACCT 587
Db 532301 GAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 532360
QY 588 GAGCAGAGCGGGCCAGGTGTGAGCTGAGATCCCGGCGCTCGTGGCAGAGACGCTGG 647
Db 532361 GAGGAGGAGGAGGAGGAGTGTGAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 532420
QY 648 GGAGTACCTGTGTGATGTGCGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 707
Db 532421 GGAGGAGATGAGGAGGAGGAGGAGTGTGAG-GAAGAGAGAGAGGAGGAGGAGGAGGAGGA 532479

Qy 708 GCCTTCCAGTTTCATAGAGGTCTGAGGAATGAAGAGCCACAGAAAGGGGACACGGCCAC 767
Db 532480 GGAGGAGGAGGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532539
Qy 768 GCTGTGTGTGAGCTGAGCAAGGCGGACCCGCTGGAGTGGAGGAAGGGGCGATGAGACCT 827
Db 532540 GGAGGAGGTGAAGAGGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532599
Qy 828 CAGAGATGGGACACAGACACCTTGAGCAGGACGGGTCCAGGTGTGAGCTGTCAGATCCG 887
Db 532600 GGAGGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532659
Qy 888 TGGCTGT 943
Db 532660 GGAGGAGGTGAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532715

RESULT 4
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (1)-(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match 4.2%; Score 42.4; DB 3; Length 1926;
Best Local Similarity 41.5%; Pred. No. 0.46;
Matches 344; Conservative 0; Mismatches 481; Indels 3; Gaps 1;
Qy 40 GATGGGACAGATACATCTTGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 99
Db 1117 GAGAGGGGACCGAGAAAGTCTGAAGGGGGGACCGAGAAAGTGCAGAGATGGGGCG 1058
Qy 100 CTGGCCATGGCGGACCGGGAGTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 159
Db 1057 CCGAGGTGATGGAGGAGCTGGGGCCGAGGTGATGGAGGAGCTGGGGCCGAGGTGACG 998
Qy 160 ACCTCACCATCAGGGCTCTGCCACAGGTTTCATAGAAGATGTGAAACACAGAGGCC 219
Db 997 GAGAGCTGGGGCCGAGGTGAGGACGAGACCGGGGAGGACGAGGACCGGGGAGGAGAC 938
Qy 220 AGAAGGGGCGCA CGGCTGT 279
Db 937 GAGGACGGGAGGACCGGGAGGACGAGGACCGGGGAGGACGAGGACCGGGGAGGAGGAC 758
Qy 397 GTGTGCGGGCAGGAGGACCTCGGCTATGTCTACCGTTCAGGCTTACCCATCAAGTTC 456
Db 757 GGGGAGGACCGGGAGGACGAGGACCGGGGAGGACGAGGACCGGGGAGGACGAGGACCGGG 698
Qy 457 ACAGAGGTCTGAGGAACAGAGGCCACAGAGGGGCAACAGCGGTCTGTGTGTGTGTGTGTGT 516

Db 697 GACGAGGACCGGGAGGAGCGGGGAGGACCGGGGAGGAGGACGAGGACCGGGGAGGACGAG 638
Qy 517 CTGAGCAAGATGGCCCCCTGGAGTGTGTGAAGGGGATGAGACCTCTAGAGATGGAGAC 576
Db 637 GGGGAGGACCGGGGAGGACCGGGGAGGACGAGGACCGGGGAGGAGGACGAGGACCGGGGAG 578
Qy 577 AGACACAGCTTGAGGACGAGGACCGGGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 636
Db 577 GGGGAGGACCGGGGAGGAGGACGAGGACCGGGGAGGACCGGGGAGGAGGACGAGGACCGGG 518
Qy 637 GAGGACCTCTGGGAGTACCTGT 696
Db 517 GAGGACGAGGACCGGGGAGGACCGGGGAGGACCGGGGAGGAGGACGAGGACCGGGGAGG 458
Qy 697 GTCAAGGCCATGCTTCCAGTTTCATAGAGGTTCTGAGGAAATGAAGAGGCCACAGAGGG 756
Db 457 GAGGAGGACGAGGACCGGGGAGGAGGACGAGGACCGGGGAGGACCGGGGAGGAGGAGGAG 398
Qy 757 GACAGGGCCACGCTGT 816
Db 397 GACGAGGACCGGGGAGGAGGACGAGGACCGGGGAGGACCGGGGAGGAGGAGGAGGAGG 338
Qy 817 CATGAGACCTCTGAGAGTGGGACAGACACAGCTGAGGCGAGGACCGG 864
Db 337 GAGGACGAGGACCGGGGAGGAGGACGAGGACCGGGGAGGACCGGGGAGGAGGAGGAG 290

RESULT 5
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassem B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match 4.2%; Score 42.4; DB 2; Length 1931;
Best Local Similarity 41.5%; Pred. No. 0.46;
Matches 344; Conservative 0; Mismatches 481; Indels 3; Gaps 1;
Qy 40 GATGGGACAGATACATCTTGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 99
Db 1117 GAGAGGGGACCGAGAAAGTCTGAAGGGGGGACCGAGAAAGTGCAGAGATGGGGCG 1058
Qy 100 CTGGCCATGGCGGACCGGGAGTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 159
Db 1057 CCGAGGTGATGGAGGAGCTGGGGCCGAGGTGATGGAGGAGCTGGGGCCGAGGTGACG 998
Qy 160 ACCTCACCATCAGGGCTCTGCCACAGGTTTCATAGAAGATGTGAAACACAGAGGCC 219
Db 997 GAGGAGCTGGGGCCGAGGTGAGGACGAGGACCGGGGAGGACGAGGACCGGGGAGGAGG 938
Qy 220 AGAAGGGGCGCA CGGCTGT 279
Db 937 GAGGACGGGAGGACCGGGGAGGACGAGGACCGGGGAGGACGAGGACCGGGGAGGAGG 878
Qy 280 AGAAGGGGCTGAGAGCCCTCAGAGATGGGACAGATACAGCTTGAGGACGAGGAGGAG 339
Db 877 GACGGGAGGACGAGGACCGGGGAGGACCGGGGAGGAGGACGAGGACCGGGGAGGAGG 818

QY 340 AAATGTGAGCTCAGATTCTGTGGCTTGGCCATG---GCAGACACTGGGGAGTACTCTGTGC 396
Db |||||
QY 817 GAGGACGAGACCGGGAGGACCGGGAGGACGAGGACCGGGAGGACCGGGAGGACGAGGAC 758
Db |||||
QY 397 GTGTGCGGCGAGGAGAGACCTCGGCTATGCTACCGTACGGCTCTACCCATCAAGTTC 456
Db |||||
QY 757 GGGGAGGACCGGGAGGACGAGGACCGGGAGGACGAGGACCGGGAGGACGAGGACCGGGAG 698
Db |||||
QY 457 ACAGAGGGGTCTGAGGAACGAGAGGCCACAGAGGGGCAACAGCCGCTCTCGGTGTGAG 516
Db |||||
QY 697 GACGAGGACCGGGAGGACCGGGAGGACCGGGAGGAGGACGAGGACCGGGAGGACGAGGAC 638
Db |||||
QY 517 CTGACAAAGTGGCCCCGTGTGAGTGTGGTGAAGGGGCATGAGACCTCAGAGATGGAGAC 576
Db |||||
QY 637 GGGGAGGACCGGGAGGACCGGGAGGACGAGGACCGGGAGGAGGACGAGGACCGGGAGGAC 578
Db |||||
QY 577 AGACACGCTGAGCAGGACCGGGCCAGGTGTGAGCTGAGATCCGCGGCTGTGTGCA 636
Db |||||
QY 577 GGGGAGGACCGGGAGGAGGACGAGGACCGGGAGGACCGGGAGGAGGACGAGGACCGGGAG 518
Db |||||
QY 637 GAGGACGCTGGGGAGTACTGTGCTATGTGCGGAAGGAGAGGACCTCAGCCATGCTCACC 696
Db |||||
QY 517 GAGGACGAGGACCGGGAGGACCGGGAGGACCGGGAGGAGGACGAGGACCGGGAGGACGAG 458
Db |||||
QY 697 GTCAGGCGCATGCTCTCAAGTTCTATAGAGGTCTGAGGAATGAAGAGGCCACAGAGGG 756
Db |||||
QY 457 GAGGAGGACGAGGACCGGGAGGAGGACGAGGACCGGGAGGAGGACCGGGAGGAGGAG 398
Db |||||
QY 757 GACAGGCGCACGCTGTGTGTGCTGAGCTGAGCAGGCGGACCGCTGAGTGGAGTGGAGAGGG 816
Db |||||
QY 397 GACGAGGACCGGGAGGAGGACGAGGACCGGGAGGACCGGGAGGAGGACCGGGAGGAG 338
Db |||||
QY 817 CATGACACCTCAGAGATGGGACAGACAGCCTGAGGACGAGGACGAGGACGAGGACGAG 864
Db |||||
QY 337 GAGGACGAGGACCGGGAGGAGGACGAGGACCGGGAGGACGAGGACGAGGACGAGGAG 290
Db |||||

RESULT 6

US-09-949-016-4900/c
; Sequence 4900, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4900
; LENGTH: 4276
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4900

Query Match 4.2%; Score 42.2; DB 4; Length 4276;
Best Local Similarity 50.8%; Pred. No. 0.65;
Matches 101; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 492 GGCAACAGCGCTGCTCGGCTGTGAGCTGAGCAAGATGGCCCCCGTGGAGTGGTGAAGGG 551
Db |||||
QY 310 GCCGGCAGCGCGCTCCCTCCACTTTGATAGCGGTGTCCACCGAGGGCTGCTCGCAGGA 251
Db |||||
QY 552 GCATGAGACCTCAGAGATGGAGACAGACAGCCTGAGGACGAGCGGGGCCAGGTGTGA 611
Db |||||
QY 250 GGCGAGGCGCGTGTGCTGCTGCGGCGCGGCGGAGGCTGCGGGCTGCGCCGA 191
Db |||||

QY 612 GCTCAGATCGCGGCCTCTGTGCAGAGGACGCTGGGAGTACTGTGCTATGTGCGGAA 671
Db 190 GGCGCCGGGGCTCTCGGGCTCGGCCCGCAGGCTTGGGGGCTCTCTGCTCGCGGGGGA 131
QY 672 GGAGAGGACCTCAGCCATG 690
Db 130 CGCGTCGCCCTCGGCGATG 112

RESULT 7

US-09-949-016-744/c
; Sequence 744, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 744
; LENGTH: 5065
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-744

Query Match 4.2%; Score 42.2; DB 4; Length 5065;
Best Local Similarity 50.8%; Pred. No. 0.68;
Matches 101; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 492 GGCAACAGCGCTGCTCGGCTGTGAGCTGAGCAAGATGGCCCCCGTGGAGTGGTGAAGGG 551
Db 1157 GCCGGCAGCGCGCTCCCTCCACTTTGATAGCGGTGTCCACCGAGGGCTGCTCGCAGGA 1098
QY 552 GCATGAGACCTCAGAGATGGAGACAGACAGCCTGAGGACGAGCGGGCCAGGTGTGA 611
Db 1097 GGCGAGGCGCGCTGCGGTGGCTGCTGGGCGCGCGCGGAGGCTGCGGGCTGCGCCGA 1038
QY 612 GCTCAGATCGCGGCCTCTGTGCAGAGGACGCTGGGGAGTACTGTGCTATGTGCGGAA 671
Db 1037 GGCGCGCGGCGCTCGGGCTCGGCCCGCAGGCTTGGGGGCGTCTCTCTCGCGGGGGA 978

QY 672 GGAGAGGACCTCAGCCATG 690
Db 977 CGCGTCGCCCTCGGCGATG 959

RESULT 8

US-09-949-016-16642/c
; Sequence 16642, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

; SOFTWARE: FASL3EQ FOR WINDOW

```

; SOFTWARE: FABLSEQ FOR WINDOWS VERSION 4.0

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:02:07 ; Search time 175.678 Seconds
(without alignment)
9323.380 Million cell updates/sec

Title: US-10-077-130-4_COPY_22500_23500
Perfect score: 1001
Sequence: 1 ggcagtcagtgacactggc.....ctgctcaagtcgtggacct 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	999.4	99.8	5207	4	US-09-858-664A-1
2	999.4	99.8	5207	4	US-10-274-978-1
3	999.4	99.8	5207	4	US-10-274-978-3
4	999.4	99.8	5207	4	US-10-697-263-1
5	999.4	99.8	5207	4	US-10-697-263-3
6	164.8	16.5	182	4	US-09-513-999C-29729
7	92	9.2	2157	4	US-09-949-016-1966
8	90.4	9.0	2105	4	US-09-949-016-195
9	90.4	9.0	2132	2	US-09-159-385-3
10	90.4	9.0	2132	3	US-09-186-277-3
11	84.2	8.4	1584	4	US-09-799-451-205
12	82.8	8.3	1282	4	US-08-878-989-12
13	82.8	8.3	1282	3	US-09-272-796-12
14	82.8	8.3	1282	4	US-09-016-434-953
15	82	8.2	1429	2	US-09-159-385-4
16	82	8.2	1429	3	US-09-186-277-4
17	78	7.8	1788	4	US-09-797-039-9
18	78	7.8	2046	4	US-09-797-039-7
19	78	7.8	15376	4	US-09-949-016-11937
20	78	7.8	15621	4	US-09-949-016-13708
21	73.2	7.3	8865	4	US-09-949-016-4238
22	73.2	7.3	8906	2	US-08-826-267-1
23	70.4	7.0	1458	4	US-09-230-896C-5
24	69.6	7.0	4935	2	US-08-631-097-3
25	69.6	7.0	5886	3	US-08-810-712-9
26	68.6	6.9	6034	4	US-09-949-016-711
27	68.6	6.9	6043	4	US-09-949-016-2582

28	58.6	6.9	6046	4	US-09-949-016-5830	Sequence 5830, Ap
29	67	6.7	2169	4	US-09-016-434-1147	Sequence 1147, Ap
30	67	6.7	7220	4	US-09-949-016-3586	Sequence 3586, Ap
31	63.4	6.3	678	4	US-09-841-683-4	Sequence 4, Appl
32	63.4	6.3	711	4	US-09-841-683-6	Sequence 6, Appl
33	63.4	6.3	1191	4	US-09-841-683-10	Sequence 10, Appl
34	63.4	6.3	1224	4	US-09-841-683-8	Sequence 8, Appl
35	63.4	6.3	1485	4	US-09-801-876B-1	Sequence 1, Appl
36	63.4	6.3	1485	4	US-10-254-869-1	Sequence 1, Appl
37	63.4	6.3	1485	4	US-10-667-442-1	Sequence 1, Appl
38	63.4	6.3	1675	4	US-09-841-683-12	Sequence 12, Appl
39	62.2	6.2	2673	3	US-09-428-711A-1	Sequence 1, Appl
40	62.2	6.2	3995	3	US-09-428-711A-13	Sequence 13, Appl
41	61.4	6.1	3228	4	US-09-579-664B-5	Sequence 5, Appl
42	61.4	6.1	3228	4	US-10-355-975A-5	Sequence 5, Appl
43	61.4	6.1	3503	4	US-09-823-038A-58	Sequence 58, Appl
44	61	6.1	2112	4	US-10-116-326-5	Sequence 5, Appl
45	61	6.1	2334	4	US-10-003-690-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-858-664A-1
; Sequence 1, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858.664A
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-858-664A-1

Query Match	99.8%;	Score	999.4;	DB 4;	Length	5207;
Best Local Similarity	99.9%;	Pred. No.	9.6e-222;			
Matches 1000;	Conservative	0;	Mismatches	1;	Indels	0;
Qy	1	GGCCAGTCAGTGACACTGGCCTGCCAGGTGTACGCCAGCCAGCAGCTGCCAGGCCACCTGG	60			
Db	3628	GGCCAGTCAGTGACACTGGCCTGCCAGGTGTACGCCAGCCAGCAGCTGCCAGGCCACCTGG	3687			
Qy	61	AGCAAAAGACGGAGCCCCCTGGAGAGCAGCAGCGGTGCTCTATCTCTGCACCCCTCAAG	120			
Db	3688	AGCAAAAGACGGAGCCCCCTGGAGAGCAGCAGCGGTGCTCTATCTCTGCCACCCCTCAAG	3747			
Qy	121	AAC TTCAGCTTCTGACCATCTCTGGTGGTGTGAGGACCTGGGTGTGTACACCTGC	180			
Db	3748	AAC TTCAGCTTCTGACCATCTCTGGTGGTGTGAGGACCTGGGTGTGTACACCTGC	3807			
Qy	181	AGCGTGAGCAATCGCGTGGGACAGTGACACACAGGGCGTCTCCGGAAGGCAGAGCGC	240			
Db	3808	AGCGTGAGCAATCGCGTGGGACAGTGACACACAGGGCGTCTCCGGAAGGCAGAGCGC	3867			
Qy	241	CCCTCATCTTTCGCCATGCCCGATATCGGGGAGGTGTACCGGATGGGGTGTCTGTCTGTC	300			
Db	3868	CCCTCATCTTTCGCCATGCCCGATATCGGGGAGGTGTACCGGATGGGGTGTCTGTCTGTC	3927			
Qy	301	TGGAAGCCCGTGGAAATCTCTACGGCCCTGTGACCTACATTGTGCAGTGCAGCTAGAAGGC	360			
Db	3928	TGGAAGCCCGTGGAAATCTCTACGGCCCTGTGACCTACATTGTGCAGTGCAGCTAGAAGGC	3987			

Qy	361	GGCAGCTGGACCACTGGCCCTCGACATCTTTGACTGTGCTACTGTGACCGACAAGTC	420
Db	3988	GGCAGCTGGACCACTGGCCCTCGACATCTTTGACTGTGCTACTGTGACCGACAAGTC	4047
Qy	421	TCCCGGGTGGCACCCTACACCTTCGCGACGGCATGTGTCACGAGGCGAGGAATGGTCCC	480
Db	4048	TCCCGGGTGGCACCCTACACCTTCGCGACGGCATGTGTCACGAGGCGAGGAATGGTCCC	4107
Qy	481	TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACCTGGCCTCTGAGGAG	540
Db	4108	TACAGCAGCCCTCGGAGCAAGTCTCTCTGGAGGGGCCAGCCACCTGGCCTCTGAGGAG	4167
Qy	541	GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTCGCATTCACAGCA	600
Db	4168	GAGAGCCAGGGCGGTGAGCCCAACCCCTGCCAGCACAAGACCTTCGCATTCACAGCA	4227
Qy	601	CAGATCCAGAGGGGCCGCTTCAGCGGTGTGCGGCAATGTCTGGGAGAAGGCCAGCGGCGG	660
Db	4228	CAGATCCAGAGGGGCCGCTTCAGCGGTGTGCGGCAATGTCTGGGAGAAGGCCAGCGGCGG	4287
Qy	661	CGCTGGCGGCCAAGATCATCCCTTACACCCCAAGGACAAGACGAGTGTCTGCGCGAA	720
Db	4288	CGCTGGCGGCCAAGATCATCCCTTACACCCCAAGGACAAGACGAGTGTCTGCGCGAA	4347
Qy	721	TACAGGCGCTCAAGGGCTTCGCGCACCCGACCTGGCCAGCTGACGACGAGCCTACCTC	780
Db	4348	TACAGGCGCTCAAGGGCTTCGCGCACCCGACCTGGCCAGCTGACGACGAGCCTACCTC	4407
Qy	781	AGCCCCGGCACCTGGTGCTCATCTTGGAGCTGTGTCTCTGGGCCCGAGCTGCTCCCTGC	840
Db	4408	AGCCCCGGCACCTGGTGCTCATCTTGGAGCTGTGTCTCTGGGCCCGAGCTGCTCCCTGC	4467
Qy	841	CTGGCCAGAGGGCCCTCTTACTCAGAATCCGAGGTGAAGGACTACTGTGGCAGATGTTG	900
Db	4468	CTGGCCAGAGGGCCCTCTTACTCAGAATCTGAGGTGAAGGACTACCTGTGGCAGATGTTG	4527
Qy	901	AGTGCCACCCAGTACCTTGCAACACGACACATCTCTGCACCTGGACCTCGAGAAC	960
Db	4528	AGTGCCACCCAGTACCTTGCAACACGACACATCTCTGCACCTGGACCTCGAGAAC	4587
Qy	961	ATGATCATCCGGAATACACCTGCTCAAGGTCTGTGACCT	1001
Db	4588	ATGATCATCCGGAATACACCTGCTCAAGGTCTGTGACCT	4628

RESULT 2

```

US-10-274-978-1
; Sequence 1, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 03/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 03/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
US-10-274-978-1

```

Qy	1	GGCCAGTCAGTGA	CAC	TG	GGCCTG	CCAGAGTGT	CAGCCAGC	AGCAGCTG	CC	CAGGCG	CACTGG	60	
Db		GGCCAGTCAGTGA	CAC	TG	GGCCTG	CCAGAGTGT	CAGCCAGC	AGCAGCTG	CC	CAGGCG	CACTGG	3687	
Qy	61	AGCAAAACGGAG	CCCCCT	TGG	AGAGAG	CAGAGCCG	GTGCTCT	ATCT	TG	CCACCC	CTCAAG	120	
Db		AGCAAAACGGAG	CCCCCT	TGG	AGAGAG	CAGAGCCG	GTGCTCT	ATCT	TG	CCACCC	CTCAAG	3747	
Qy	121	AAC	TTCCAG	CTT	CTG	ACCAT	CC	TG	TG	TGGT	GGCT	TGAGG	180
Db		AAC	TTCCAG	CTT	CTG	ACCAT	CC	TG	TG	TGGT	GGCT	TGAGG	3807
Qy	181	AGCGTAGCA	AA	TG	CGCT	TGGGA	CAG	TGA	CA	CA	CGGCGT	CT	240
Db		AGCGTAGCA	AA	TG	CGCT	TGGGA	CAG	TGA	CA	CA	CGGCGT	CT	3867
Qy	241	CCCT	CA	TCT	TG	CGCAT	CGG	ATAT	CG	GGGAG	GGT	GTAC	300
Db		CCCT	CA	TCT	TG	CGCAT	CGG	ATAT	CG	GGGAG	GGT	GTAC	3927
Qy	301	TGGA	AG	CCCG	TG	GAA	TCT	TAC	GGC	CCCT	GTG	AC	360
Db		TGGA	AG	CCCG	TG	GAA	TCT	TAC	GGC	CCCT	GTG	AC	3987
Qy	361	GGC	AG	CT	TG	GAC	CA	CT	TG	TG	AGT	CT	420
Db		GGC	AG	CT	TG	GAC	CA	CT	TG	TG	AGT	CT	4047
Qy	421	TCC	CG	GG	GGT	GG	CAC	T	T	CCG	CA	T	480
Db		TCC	CG	GG	GGT	GG	CAC	T	T	CCG	CA	T	4107
Qy	481	TAC	G	AG	C	AG	CC	CT	T	CG	CA	A	540
Db		TAC	G	AG	C	AG	CC	CT	T	CG	CA	A	4167
Qy	541	GAG	AG	CC	AG	GGG	GG	GT	C	AG	CC	CA	600
Db		GAG	AG	CC	AG	GGG	GG	GT	C	AG	CC	CA	4227
Qy	601	CAG	AT	CA	G	AG	GG	CG	CG	CT	T	T	660
Db		CAG	AT	CA	G	AG	GG	CG	CG	CT	T	T	4287
Qy	661	GC	CT	TG	CG	CG	C	NA	G	AT	C	A	720
Db		GC	CT	TG	CG	CG	C	NA	G	AT	C	A	4347
Qy	721	TAC	G	AG	CG	CC	T	CA	A	AG	CG	CG	780
Db		TAC	G	AG	CG	CC	T	CA	A	AG	CG	CG	4407
Qy	781	AG	CC	CC	CG	G	A	C	T	G	T	G	840
Db		AG	CC	CC	CG	G	A	C	T	G	T	G	4467
Qy	841	CT	G	CG	CG	AG	GG	CC	T	CT	A	C	900
Db		CT	G	CG	CG	AG	GG	CC	T	CT	A	C	4527
Qy	901	AG	T	G	CC	AC	CC	AG	T	A	C	T	960
Db		AG	T	G	CC	AC	CC	AG	T	A	C	T	4587
Qy	961	AT	G	A	T	C	A	T	C	C	G	A	1001
Db		AT	G	A	T	C	A	T	C	C	G	A	4628

RESULT 3

US-10-274-978-3
; Sequence 3, Application US/10274978
; Patent No. 6670164


```
US-09-513-999C-29729
; Sequence 29729, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 29729
; LENGTH: 182
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 139
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 152
; OTHER INFORMATION: y=c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 154
; OTHER INFORMATION: y=c or t
; OTHER INFORMATION: y=c or t
US-09-513-999C-29729

Query Match      16.5%; Score 164.8; DB 4; Length 182;
Best Local Similarity 97.8%; Pred. No. 5.5e-29;
Matches 174; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Qy 613 GCGCGCTTACGCTGTCGCGCAATGCTGGAGAGGCGAGCGCGCGCGCTGCGCGCC 672
Db 1 GCGCGCTTACGCTGTCGCGCAATGCTGGAGAGGCGAGCGCGCGCGCTGCGCGCC 60

Qy 673 AAGATCATCCCTACCCACCCCAAGACAGCAGTGTGCGGCAATACGAGCGCCCTC 732
Db 61 AAGATCATCCCTACCCACCCCAAGACAGCAGTGTGCGGCAATACGAGCGCCCTC 120

Qy 733 AAGGCGCTGCGCCACCCGACCTGCGCCAGCTGCA-CGAGCTACCTCAGCCCGCGG 789
Db 121 AAGGCGCTGCGCCACCCGACCTGCGCCAGCTGCA-CGAGCTACCTCAGCCCGCGG 178

RESULT 7
US-09-949-016-1966
; Sequence 1966, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1966
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1966

Query Match      9.0%; Score 90.4; DB 4; Length 2105;
Best Local Similarity 58.0%; Pred. No. 1.8e-11;
Matches 160; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 714 GCGGAATACAGAGCGCCCTCAAGGGCCTGCGCCACCCGACCTGCGCCAGCTGCGCAGC 773
Db 279 GCGGAGGTGAACATCCTCGCGGAGATCCGCGACCCCAACATCATCACCTGCGCAGCAT 338

Qy 774 CTACTCAGCCCCCGGACCTGCTGCTCATCTTGAGCTGTGCTCTGCGCCGAGCTGCT 833
Db 339 CTTCGAGAACAGACGCGAGCTGCTCATCTCGAGCTGTGCTCTGCGCGGGAGCTCTT 398

Qy 834 CCCCTGCTGCGCGAGAGGGCTCTCTACTAGAAATCCGAGGTGAAGGACTACCTGTGCA 893
Db 399 TGACTTCTCGCGGAGAAAGAGTCTGACGAGAGACGAGGCCACCCAGTTCTCTCAAGCA 458

Qy 894 GATGTTGAGTGCCACCCAGTACCTGCGCAACACGACCATCTCGACCTGGACCTGAGGTC 953
Db 459 GATCTGAGCGCGCTTCACTACTGCTGCACTCTAAGGCGATCGCACACTTTGACCTGAAGCC 518

; ORGANISM: Human
US-09-949-016-1966

Query Match      9.2%; Score 92; DB 4; Length 2157;
Best Local Similarity 58.3%; Pred. No. 7.6e-12;
Matches 161; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Qy 714 GCGGAATACAGAGCGCCCTCAAGGGCCTGCGCCACCCGACCTGCGCCAGCTGCGCAGC 773
Db 331 GCGGAGGTGAACATCCTCGCGGAGATCCGCGACCCCAACATCATCACCTGCGCAGCAT 390

Qy 774 CTACTCAGCCCCCGGACCTGCTGCTCATCTTGAGCTGTGCTCTGCGCCGAGCTGCT 833
Db 391 CTTCGAGAACAGACGCGAGCTGCTCATCTCGAGCTGTGCTCTGCGCGGGAGCTCTT 450

Qy 834 CCCCTGCTGCGCGAGAGGGCTCTCTACTAGAAATCCGAGGTGAAGGACTACCTGTGCA 893
Db 451 TGACTTCTCGCGGAGAGAGTCTGACGAGAGACGAGGCCACCCAGTTCTCTCAAGCA 510

Qy 894 GATGTTGAGTGCCACCCAGTACCTGCGCAACACGACCATCTCGACCTGGACCTGAGGTC 953
Db 511 GATCTGAGCGCGCTTCACTACTGCTGCACTCTAAGGCGATCGCACACTTTGACCTGAAGCC 570

Qy 954 CGAGAACATGATCATCACCGAATACACCTGCTCAA 989
Db 571 GGAAGACATCATGCTGCTGGACAAGAGCTGCCCAA 606

RESULT 8
US-09-949-016-195
; Sequence 195, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195
; LENGTH: 2105
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195
```

QY 954 CGAGAACATGATCATCACCGAATAACAACCTGCTCAA 989
Db 519 GGAATAACATCATGCTGCTGGACAAGACGTGCCCAA 554

RESULT 9

US-09-159-385-3
; Sequence 3, Application US/09159385
; Patent No. 5958748
; GENERAL INFORMATION:
; APPLICANT: AKIRA, SHIZUO
; APPLICANT: KAWAI, TARO
; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
; FILE REFERENCE: PH-569
; CURRENT APPLICATION NUMBER: US/09/159,385
; CURRENT FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: JP97/261589
; EARLIER FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2132
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (94)..(1455)
US-09-159-385-3

Query Match 9.0%; Score 90.4; DB 2; Length 2132;
Best Local Similarity 58.0%; Pred. No. 1.8e-11;
Matches 160; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 714 GCGGAATACGAGGCCCTCAAGGGCCTGGCCACCCGACCTGGCCAGCTGCGACGACG 773
Db 279 GCGGAGGTGAACATCTCTGGGAGATCCGGACCCCAACATCATCACCTGCGACGACAT 338
QY 774 CTACTCAGCCCCCGGACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCCCGGAGCTGCT 833
Db 339 CTTCGAGAACAGACGGAGCTGGTCTCATCTCTGGAGCTGGTCTCTGGCGGGAGCTCTT 398
QY 834 CCCCTGCTGGCGAGAGGGCTCTCTACTCAGATCCGAGTGAAGGACTACCTGTGGCA 893
Db 399 TGACTTCTGGCGGAGAAAGATCGCTGACGGAGGACGAGGCCACCCAGTTCTCTCAAGCA 458
QY 894 GATGTTGAGTGCACCCAGTACCTGCACACACGACATCTGTGACCTGGACCTGAGGTC 953
Db 459 GATCTGACGGGGTTCATCTGCTGCTCTTAAGCGCATCGCACACTTTGACCTGAAGCC 518
QY 954 CGAGAACATGATCATCACCGAATAACAACCTGCTCAA 989
Db 519 GGAATAACATCATGCTGCTGGACAAGACGTGCCCAA 554

RESULT 10

US-09-186-277-3
; Sequence 3, Application US/09186277
; Patent No. 6171841
; GENERAL INFORMATION:
; APPLICANT: AKIRA, SHIZUO
; APPLICANT: KAWAI, TARO
; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
; FILE REFERENCE: 081356/0128
; CURRENT APPLICATION NUMBER: US/09/186,277
; CURRENT FILING DATE: 1998-11-05
; EARLIER APPLICATION NUMBER: JP97/261589
; EARLIER FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2132
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (94)..(1455)
US-09-186-277-3

Query Match 9.0%; Score 90.4; DB 3; Length 2132;

Best Local Similarity 58.0%; Pred. No. 1.8e-11;
Matches 160; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 714 GCGGAATACGAGGCCCTCAAGGGCCTGGCCACCCGACCTGGCCAGCTGCGACGACG 773
Db 279 GCGGAGGTGAACATCTCTGGGAGATCCGGACCCCAACATCATCACCTGCGACGACAT 338
QY 774 CTACTCAGCCCCCGGACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCCCGGAGCTGCT 833
Db 339 CTTCGAGAACAGACGGAGCTGGTCTCATCTCTGGAGCTGGTCTCTGGCGGGAGCTCTT 398
QY 834 CCCCTGCTGGCGAGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTGGCA 893
Db 399 TGACTTCTGGCGGAGAAAGATCGCTGACGGAGGACGAGGCCACCCAGTTCTCTCAAGCA 458
QY 894 GATGTTGAGTGCACCCAGTACCTGCACACACGACATCTGTGACCTGGACCTGAGGTC 953
Db 459 GATCTGACGGGGTTCATCTGCTGCTCTTAAGCGCATCGCACACTTTGACCTGAAGCC 518
QY 954 CGAGAACATGATCATCACCGAATAACAACCTGCTCAA 989
Db 519 GGAATAACATCATGCTGCTGGACAAGACGTGCCCAA 554

RESULT 11

US-09-799-451-205
; Sequence 205, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Aeundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yungqing
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Dunrui
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6783969el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803
; CURRENT APPLICATION NUMBER: US/09/799,451
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: Pt_FL_genes Version 2.0
; SEQ ID NO 205
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (51)..(1166)
US-09-799-451-205

Query Match 8.4%; Score 84.2; DB 4; Length 1584;
Best Local Similarity 54.3%; Pred. No. 4.5e-10;
Matches 195; Conservative 0; Mismatches 158; Indels 6; Gaps 1;

NAME: Billings, Lucy J J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0321 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNOT06
CLONE: 827431
US-09-272-796-12

Query Match 8.3%; Score 82.8; DB 3; Length 1282;
Best Local Similarity 52.5%; Pred. No. 8.9e-10;
Matches 208; Conservative 0; Mismatches 182; Indels 6; Gaps 1;

QY 573 CAGCAAAAGACCTTCGCATTCCAGACACAGATCCAGAGGGCGCGCTTCAGGGTGGTGG 632
DB 230 CATCAGCAGCGTCTACGAGATCCGAGAGGCTCGGCTCGGCTGCTTCTCCGAGGTGGT 289

QY 633 GCAATGCTGGGAGAGGCGGCGGCGCTGCGCGCCAGATCATCCCTTA----- 686
DB 290 GCTGGCCAGGAGCGGGCTCGCACACCTCGTGGCCCTCAAGTCATCCCAAGAGGC 349

QY 687 CCACCCCAAGGACAGACAGAGTCTGCGGGAATACAGAGGCCCTCAAGGGCTCGGCCA 746
DB 350 CCTCGGGGCAAGGAGGCGCTGTTGAGAACGAGATCGAGTGTCTCGTAGGATCAGTCA 409

QY 747 CCGGCACTGGCCAGCTGCAAGCAGCTTACCTCAGCCCGCGGCGGCTGCTCATCTT 806
DB 410 CCCCACATCTGCTGAGGATGTCCAGAGAGCCCTTCCACCTCTACTTGGCCAT 469

QY 807 GGAGCTGTGCTTGGGCGGAGTCTCCCTGCTGCGGAGAGGGCTCTCTACTCAGA 866
DB 470 GGAATGCTGAGCGGTGGCGAGCTGTTGACCGCATCATGAGGCGGCTCTTACACAGA 529

QY 867 ATCCAGGTGAGGACTACCTGTGCGAGATGTTGAGTGCACCCAGTACCTGCACACCA 926
DB 530 GAAGATGCGAGCCATCTGTGGTGGTCAAGTCTTGGCGCGCTTCTTCTACTGCAAGCCT 589

QY 927 GCACATCTGCACCTGGACCTGAGTCCGAGAACAT 962
DB 590 GGGGATCTGCACCGGAGCTCAAGCCCGAAACCT 625

RESULT 14
US-09-016-434-953
Sequence 953, Application US/09016434
Patent No. 6500938
GENERAL INFORMATION:
APPLICANT: Janice Au-Young
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
TITLE OF INVENTION: PATHWAY GENE EXPRESSION
NUMBER OF SEQUENCES: 1490
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 953:
SEQUENCE CHARACTERISTICS:
LENGTH: 1282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNOT06
CLONE: 827431
US-09-016-434-953

Query Match 8.3%; Score 82.8; DB 4; Length 1282;
Best Local Similarity 52.5%; Pred. No. 8.9e-10;
Matches 208; Conservative 0; Mismatches 182; Indels 6; Gaps 1;

QY 573 CAGCAAAAGACCTTCGCATTCCAGACACAGATCCAGAGGGCGCGCTTCAGGGTGGTGG 632
DB 230 CATCAGCAGCGTCTACGAGATCCGAGAGGCTCGGCTCGGCTGCTTCTCCGAGGTGGT 289

QY 633 GCAATGCTGGGAGAGGCGGCGGCGCTGCGCGCCAGATCATCCCTTA----- 686
DB 290 GCTGGCCAGGAGCGGGCTCCCGCACACCTGCTGGCCCTCAAGTCATCCCAAGAGGC 349

QY 687 CCACCCCAAGGACAGACAGTGTGCTGCGGAATACAGAGGCCCTCAAGGGCTCGGCCA 746
DB 350 CCTCGGGGCAAGGAGGCGCTGTTGAGAACGAGATCGAGTGTCTCGTAGGATCAGTCA 409

QY 747 CCGGCACTGGCCAGCTGCAAGCAGCTTACCTCAGCCCGCGGCGGCTGCTCATCTT 806
DB 410 CCCCACATCTGCTGAGGATGTCCAGAGAGCCCTTCCACCTCTACTTGGCCAT 469

QY 807 GGAGCTGTGCTTGGGCGGAGTCTCCCTGCTGCGGAGAGGGCTCTCTACTCAGA 866
DB 470 GGAATGCTGAGCGGTGGCGAGCTGTTGACCGCATCATGAGGCGGCTCTTACACAGA 529

QY 867 ATCCAGGTGAGGACTACCTGTGCGAGATGTTGAGTGCACCCAGTACCTGCACACCA 926
DB 530 GAAGATGCGAGCCATCTGTGGTGGTCAAGTCTTGGCGCGCTTCTTCTACTGCAAGCCT 589

QY 927 GCACATCTGCACCTGGACCTGAGTCCGAGAACAT 962
DB 590 GGGGATCTGCACCGGAGCTCAAGCCCGAAACCT 625

RESULT 15
US-09-159-385-4
Sequence 4, Application US/09159385
Patent No. 5958748
GENERAL INFORMATION:
APPLICANT: AKIRA, SHIZUO
APPLICANT: KAWAI, TARO
TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
FILE REFERENCE: PH-569
CURRENT APPLICATION NUMBER: US/09/159,385
CURRENT FILING DATE: 1998-09-23
EARLIER APPLICATION NUMBER: JP97/261589
EARLIER FILING DATE: 1997-09-26
NUMBER OF SEQ ID NOS: 8


```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1429
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)..(1353)
US-09-159-385-4
```

```
Query Match      8.2%; Score 82; DB 2; Length 1429;
Best Local Similarity 58.7%; Pred. No. 1.4e-09;
Matches 142; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy      740 TCGGCCACCCGACCTGGCCCGAGCTGCGAGCGCCTACCTAGCCCCCGGACCTGGTGC 799
Db      221 TCCGCCACCCCAACATCATACACTGTCATGACGTGTCGAGAACAGACAGATGTGTGC 280

Qy      800 TCATCTTGGAGTGTGCTCTGGGCCGAGCTGCTCCCTGCTGGCCGAGAGGGCCTCT 859
Db      281 TGATCCTGGAGTGTGTGTCGGTGGCGAGCTTTTCGACTTCTGGCCGAGAGGAGTCAT 340

Qy      860 ACTCAGAAATCCGAGGTGAAGGACTACCTGTGGCAGATGTTGAGTGCCACCCAGTACCTGC 919
Db      341 TGACGGAGGATGAGGCCACGCAGTTCTCAACAAATCCTAGACGGTGTCCACTACCTGC 400

Qy      920 ACAACCCAGACATCTGTGACCTGGACCTGAGTCCGAGAACATGATCATCACCGAATACA 979
Db      401 ACTCCAGCGCATCGCACACTTTGACCTGAAGCCCGAGAACATCATGTTGCTGGACAAGC 460

Qy      980 AC 981
Db      461 AC 462
```

Search completed: March 21, 2005, 14:38:50
Job time : 178.678 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:02:07 ; Search time 175.678 Seconds
(without alignment)
9323.380 Million cell updates/sec

Title: US-10-077-130-6_COPY_2000_3000
Perfect score: 1001
Sequence: 1 cgcctaccatccggagtg.....ctgaggcaggcccaatgcc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	830.8	83.0	2768	4	US-09-774-528-426
C 2	64	6.4	1926	3	US-09-249-585A-4
C 3	64	6.4	1931	2	US-09-130-114-2
4	62	6.2	1651	4	US-09-902-540-3255
5	62	6.2	19954	4	US-09-902-540-1150
6	52.8	5.3	2695	4	US-09-620-312D-473
7	51.6	5.2	3575	4	US-09-023-655-1395
C 8	51.4	5.1	690	4	US-09-902-540-7374
C 9	51.4	5.1	2855	4	US-09-902-540-1935
10	51.4	5.1	18316	4	US-09-902-540-1057
C 11	51	5.1	7218	1	US-08-232-463-14
12	49.4	4.9	3766	4	US-09-949-016-1651
13	49.4	4.9	3778	4	US-09-949-016-359
14	49.4	4.9	50937	3	US-09-428-517-1
15	49	4.9	2670	4	US-09-902-540-8525
C 16	49	4.9	9321	4	US-09-902-540-898
17	48.4	4.8	4285	3	US-09-040-774-1
18	48.2	4.8	1539	3	US-08-582-740-69
19	48.2	4.8	1539	3	US-09-109-879-69
20	48.2	4.8	1598	4	US-09-023-655-915
21	48.2	4.8	1624	3	US-08-582-740-67
22	48.2	4.8	1624	3	US-09-109-879-67
23	48	4.8	5207	4	US-09-858-664A-1
24	48	4.8	5207	4	US-10-274-978-1
25	48	4.8	5207	4	US-10-274-978-3
26	48	4.8	5207	4	US-10-697-263-1
27	48	4.8	5207	4	US-10-697-263-3

28	47.8	4.8	2668	2	US-08-461-775-11	Sequence 11, Appl
29	47.8	4.8	2668	3	US-09-031-606-11	Sequence 11, Appl
30	47.4	4.7	3397	4	US-09-949-016-2188	Sequence 2188, Ap
31	47	4.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
32	47	4.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
33	46.8	4.7	2199	4	US-09-902-540-6388	Sequence 6388, Ap
34	46.8	4.7	2585	4	US-09-902-540-424	Sequence 424, App
35	46.4	4.6	1320	4	US-09-252-991A-165	Sequence 165, App
36	46.4	4.6	1686	4	US-09-252-991A-155	Sequence 155, App
C 37	46.4	4.6	1803	4	US-09-252-991A-170	Sequence 170, App
C 38	46.4	4.6	2610	4	US-09-252-991A-175	Sequence 175, App
39	46.2	4.6	1620	2	US-08-461-775-10	Sequence 10, Appl
40	46.2	4.6	1620	3	US-09-031-606-10	Sequence 10, Appl
41	45.8	4.6	603	4	US-09-902-540-9070	Sequence 9070, Ap
42	45.8	4.6	7719	4	US-09-902-540-969	Sequence 969, App
C 43	45.8	4.6	77536	4	US-09-410-551B-1	Sequence 1, Appli
C 44	45.8	4.6	77536	4	US-09-940-316B-1	Sequence 1, Appli
45	45.6	4.6	3114	4	US-09-902-540-2501	Sequence 2501, Ap

ALIGNMENTS

RESULT 1

US-09-774-528-426
; Sequence 426, Application US/09774528
; Patent No. 6743619
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Yang, Yonghong
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6743619el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802
; CURRENT APPLICATION NUMBER: US/09/774,528
; CURRENT FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt_Fl_genes Version 2.0
; SEQ ID NO 426
; LENGTH: 2768
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72)..(2739)
US-09-774-528-426

Query Match		83.0%	Score 830.8;	DB 4;	Length 2768;
Best Local Similarity		95.3%	Pred. No. 4.9e-173;		
Matches		856;	Conservative	0;	Mismatches 42;
				Indels	0;
				Gaps	0;
Qy	1	CGCTCACCATCCGGAGGTGCCCCCAGCTGCGACGGGGCGCAGCTGAAGTTCGTGGCCA	60		
Db	1861	CGCTCACCATCCGGAGGTGCCCCCAGCTGCGACGGGGCGCAGCTGAAGTTCGTGGCCA	1920		
Qy	61	ACGGCATTGAGCAGCATCCGGATGAGGTCCCGGGCGGGCCCGGCTGACTGCCACA	120		
Db	1921	ACGGCATTGAGCAGCATCCGGATGAGGTCCCGGGCGGGCCCGGCTGACTGCCACA	1980		
Qy	121	AGCCCGCAGCGCAGCTGCCCGGAGGTGCTGGCTCGCTCAGCAGGAGCGCAGCTGC	180		
Db	1981	AGCCCGCAGCGCAGCTGCCCGGAGGTGCTGGCTCGCTCAGCAGGAGCGCAGCTGC	2040		


```

; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/10903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

```

Query Match	6.4%	Score 64	DB 2	Length 1931
Best Local Similarity	43.9%	Pred. No. 7.1e-05		
Matches 387	Conservative 0	Mismatches 480	Indels 15	Gaps 2
Qy	14	GGAGGTGCCCGCCAGCCTGCACCGGGCGCAGCTGAAGTTCTGTGGCCCAACGGCATTGAGAG	73	
Db	1127	GCAGGTCTCGAGAGGGGGACCGAGAAAGTGTCTAGGGGGGGACCCAGAAAGTGTGACGGA	1068	
Qy	74	CAGCATCCGGATGGAGTCCGGGCGGCGCCCGAGGCTGACTGCCAAACAAGCCCGCCAGCCGC	133	
Db	1067	GATGGGCGCCCGGAGGTGATGAGGAGCTGGGGCCGAGGTGATGAGGAGCTGGGGCC	1008	
Qy	134	AGTGTCCCGGAGGTGTGTGCTCGGTGTGCACGAGAGCGCAGCTGTGCTGTGAGCTGTCT	193	
Db	1007	GGAGGTACGCGAGCGAGCTCGGGCCGAGGTGGAGGACGAGACGGGAGGACGAGGACGG	948	
Qy	194	AGATCAGGCTCGGGCTGTGACGTGGCTGAAGGATGCTCGCACACTGTCTCCAGGCCCCCAA	253	
Db	947	GSAGGAGGACGAGGACCGGGAGGACCGGGGAGGACGAGGACCGGGAGGACCGGGAGGACGA	888	
Qy	254	GTATGAGTGCAGGCATCGCGCGGGCGGGTGCTCTTTGTGCGAGATGTGGCCCGGGA	313	
Db	887	GGACGGGAGGACGCGGGAGGACGAGGACCGGGAGGACGCGGGAGGACGAGGACCGGGAG	828	
Qy	314	CGATGCAGGCCTCTTACAGTTCGCTCAGCCCGGGGGCCGATCGCCTTACCAGTCTCTCCGT	373	
Db	827	GGACGGGGAGGAGGACGAGGACCGGGAGGACCGGGAGGACGAGGACCGGGAGGACCGGGAG	768	
Qy	374	GCAAGGCTCGCGGCTTTCTGCACAGGACATGGCGGGCAGCTGTGTGTGATGCCCTGGC	433	
Db	767	GGACGAGGACGGGG-----AGGACGGGGAGGACGAGGACCGGGAGGACGAGGACCGGGAG	720	
Qy	434	TGGGGGCCCGCGCAGTTTGAGTGTGAGACTCCGAAAGCCACCGTCCACCTGTGCACTGGTA	493	
Db	719	CGGGGAGGACGAGGACCGGGAGGACGAGGACCGGGAGGACGCGGGAGGACCGGGAGGAGGA	660	
Qy	494	CAAGGATGGCATGG---AGCTGGGCCCACTCCGCTGAGCGCTTTCTTTCAGGAGGATGTGGG	550	
Db	659	CGAGGACGGGGAGGACGAGGACCGGGAGGACCGGGAGGACCGGGAGGACGAGGACCGGGAG	600	
Qy	551	GACCGGCAACGGCTGTGTGGCAGCCACAGTCAACAGGCAGGATGAAAGCACCCTACTCTCTG	610	
Db	599	GGAGGACGAGGACCGGGAGGACCGGGAGGACCGGGAGGAGGACGAGGACCGGGAGGACCGG	540	
Qy	611	CCGCGTGGGCGAGGACTCTGTGACTTCCGCTCCGCTCTCTAGGCCCAAGGTGTGTGT	670	
Db	539	GGAGGAGGACGAGGACCGGGAGGAGGACGAGGACCGGGAGGACCGGGAGGACCGGGAGGAG	480	
Qy	671	TGCTAAGGAGCAGCTGGCACAACGAGGAAGCTGTGAGGCAGAGGACGAGGCCACCTGACACT	730	
Db	479	GGACGAGGACCGGGAGGACCGGGAGGAGGACGAGGACCGGGAGGAGGACGAGGACCGGGAG	420	
Qy	731	GAGCTGCAGGTGGCCACGCGCCNAGACGAGAGGTGACGTGGTACAAGGATGGGAAGAAGCT	790	
Db	419	GGACCGGGAGGACCGGGAGGAGGACGAGGACCGGGAGGAGGACGAGGACCGGGAGGACCGG	360	
Qy	791	GAGCTCCAGCTCGAAAGTGTGCATGGAGGCCACAGGCTGACCGCGCAGGCTGTGTGTGCA	850	
Db	359	GGAGGACCGGGAGGACCGGGAGGAGGACGAGGACCGGGAGGAGGACGAGGACCGGGAGGAG	300	
Qy	851	GCAGGACGGCCAGGCGGATGCCGGGGAGTATAGCTGCGAGGC	892	

```

Db      299   CGAGGACGGGGCGGCAGGACGAGGACGAGCAAGGTGCC    358
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 4
US-09-902-540-3255
; Sequence 3255, Application US/09902540
; Patent NO. 683347
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3255
; LENGTH: 1651
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3255
```

Query Match	6.2%	Score 62;	DB 4;	Length 1651;
Best Local Similarity	48.1%;	Pred. No. 0.00019;		
Matches 176;	Conservative 0;	Mismatches 190;	Indels 0;	Gaps 0;
Qy	318	GCAGGCCTCACAGATGCGTCAGCCGGGGGGCGCATCGCTTACCAGCTCTCCGTGCAA	377	
Db	1211	CGCGCGCTCGAAGGGCGTCTCCTTGGCGGGGGTGCCTTACATCCGCTGCTCAAG	1270	
Qy	378	GGCCTCGCGCGCTTTCTGCACAAAGGACATGGCGGGCAGCTGTGTGGATGCCGTGGCTGGG	437	
Db	1271	CGCGTGCAGCGCCTGCAGCTGTCCGGCGGTGAGAAAGTTCGGCGTGGACATCATCCGCCGC	1330	
Qy	438	GGCCCGCGCAGTTTGAGTGTGAGACCTCCGAAGCCACGTCCACAGTGCACCTGATCAAG	497	
Db	1331	GCCTGTGAGAGGCCCTCCGCAGATGTCGCGCAACGGCGGCTGGAGGACAGCTGGTG	1390	
Qy	498	GATGGCATGAGCTGGGCCCATCCCGGTGAGCGCTTCTTGACGAGGATGTGGGGACGCGG	557	
Db	1391	GTGAACAAGGCTCAAGGAGTCTTCGGGTCCGCTTCAAACGCCCGACGGGCACCTAC	1450	
Qy	558	CACCGGCTGTGGGACGCACAGTCAACGACGAGATGAAGGCACCTACTCTCTGCCGGTG	617	
Db	1451	GAGGACCTGTGTGCCCGCGCGTCATGACCCCGCCAAAGTTCAGCCGCACCCGCGCTGCAG	1510	
Qy	618	GGCGAGGACTCTGTGGACTTCCGGCTCCCGGCTCTCTGAGCCCCAAGGTGGTGTGTTGCTAAG	677	
Db	1511	AACGCGGCTCCGTGTCTCTCGCTGATGCTTACCACCGAGGCCATGTGGCGGAGCGTCCG	1570	
Qy	678	GAGCAG	683	
Db	1571	AAGGAG	1576	

RESULT 5
US-09-902-540-1150
; Sequence 1150. Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883

```

; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1150
; LENGTH: 19954
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1150

Query Match
Best Local Similarity 6.2%; Score 62; DB 4; Length 19954;
Matches 176; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

Qy 318 GCAGGCTCTACAGTGGCTCAGCGCGGGCGCGATCGCTACAGCTCTCGTGCAA 377
Db 8811 GCGGCGCTGAAGAGGCGTGTCTCGCGCGCGGTGCTCATCGCTGCCTCAAG 8870
Qy 378 GGCCTCGCGGCTTTCTGCACAAAGACATGCGGGCAGCTGTGTGATGCGGTGGG 437
Db 8871 GCGCTGAGCGGCTGCGAGCTGCGGCGGTGAGAAAGTTTCGGGTGGACATCATCGCGCG 8930
Qy 438 GGCCTCGCGGCTTTGAGTGTGAGACCTCGAAGCCAGCTCCAGCTGATGTACAG 497
Db 8931 GCGCTGAGAGGCGCTTCGCGAGATGTCGGCAACGCGGCGCTGGAGGGCAGCGTGGT 8990
Qy 498 GATGGCATGAGTGGGCACTCCGCTGAGCGCTTTCTTCGAGGAGGATGTGGGACGCG 557
Db 8991 GTGAACAAGGTCAAGAGTCTCTCGGTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 9050
Qy 558 CACGGCTGTGGCAGCAGCTACAGGAGGATGAGGACCTACTCTCTCGCGGTG 617
Db 9051 GAGGACCTCTGGCGCGCGGTGTCATGACCGCGCAAGGTGAGCGCACCGCTGCAG 9110
Qy 618 GCGGAGGACTGTGTGACTTCGCGCTCCGCGTCTCTGAGGCCAAGGTGCTTTGCTAAG 677
Db 9111 AAGCGGCGTCTGCTGCTGCTGATGTCTACACCGAGGCGTGTGGGAGCGTCCG 9170
Qy 678 GAGCAG 683
Db 9171 AAGGAG 9176

RESULT 6
US-09-620-312D-473
; Sequence 473, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yidonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radjoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784C1P2B
; CURRENT APPLICATION NUMBER: US/09/620.312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105

; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 473
; LENGTH: 2695
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (256)..(2253)
; NAME/KEY: misc.feature
; LOCATION: (1)..(2695)
; OTHER INFORMATION: n = a,t,c or g
US-09-620-312D-473

Query Match
Best Local Similarity 5.3%; Score 52.8; DB 4; Length 2695;
Matches 216; Conservative 0; Mismatches 272; Indels 0; Gaps 0;

Qy 294 GTGCGAGATGTGGCGGAGACATGAGGCTCTACGAGTGGTCAAGCGCTTCTTGTGCA 353
Db 544 GTGCGCTACAGGGCGCGCGAGGCGGTGTCAGCAGTATCAACACCTCGCCCTCGTTC 603
Qy 354 ATGCGCTACAGCTCTCCGTGCAAGGCTTCGCGCGCTTCTTGTGCAACAGGACATGG 413
Db 604 CGCGCATCCAGGACGACTGCCAGTTCATCGGCCCGCTGGCCACAGCTCGGCGAG 663
Qy 414 AGCTGTGTGATGCGGTGGCTGGGGCGCGGCGAGTTTGTGATGTGAGACCTCCGAGCC 473
Db 664 CGCTTTAGGAGGGCGGCTCAGGCGCGCGGAGCAGAGTGTGAGTGTGAGTGTGCTG 723
Qy 474 CACGTCCAGCTGCTGTGTACAAGATGGCATGGAGTGGGCTTCTTGTGGCGAGCGCTTC 533
Db 724 GCGCTGGGCGGCTTCGGAGGAGCTGTGCGAGGAGTTCTTGGCGCAGCCCGCGCGCG 783
Qy 534 TTGCGAGGAGATGTGGGAGACGCGCACCGGCTGTGTGGCAGCCAGTCCAGGACAGAT 593
Db 784 CTGGAAGAGGAGCTGAGAAACCTTGAGGCGCGAGTGGGGCGCTCACCTCCGCTCCG 843
Qy 594 GAAGGCACCTACTCTCGCGCGTGGGCGAGGACTCTGTGGACTTCCGGCTCCGCTCTCT 653
Db 844 GTGTAGAGTTTACCGACCATGGAGGAGTGTGTGGCGGCGCTCTGCGAGTGGCG 903
Qy 654 GAGCCCAAGGTGTGTGTGTAAAGAGCAGCTGGCACAGGAAAGTTCAGGCGAGAGCA 713
Db 904 GCGGCTTACAGGAGCTGTGTGGCGCGCGAGGCGCGCGAGGAGTGGCGAGCTGGCG 963
Qy 714 GGAGCCAGTGCACACTCAGCTGCGAGTGGCGCGCGCGCGAGGAGTGGAGTGGCTAC 773
Db 964 TTCGCCGCGAGCTGGGCGAGCGCTATTTTCGCTGTGTGGAGCGGCGCTGGCGAG 1023
Qy 774 AAGGATGG 781
Db 1024 CAGGCTGG 1031

RESULT 7
US-09-023-655-1395
; Sequence 1395, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESS: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1395:
SEQUENCE CHARACTERISTICS:
LENGTH: 3575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g402646
US-09-023-655-1395

```

Query Match	5.2%	Score 51.6;	DB 4;	Length 3575;
Best Local Similarity	51.8%;	Pred. No. 0.042;		
Matches 117;	Conservative 0;	Mismatches 109;	Indels 0;	Gaps 0;
QY	676	AGGAGCAGCTGGCCACGCGAGGAAGCTGCAGGCACAGAGCGCAGGACGCAGTGCACACTGAGCT	735	
DB	1338	AGGTCCTGCAGGACATCGCGGATCTGACGCTGAAGGCTCAGAACAAGCTGTGTTCAAGT	1397	
QY	736	GCAGGTGGCCCGAGCCACAGCGAGGTGACGTGGTACAAGGATGGGAAGAAGCTGAGCT	795	
DB	1398	GCAGGTGTCTGATGAGAAAGTACGCGGCAAGTGGTATAAGATGGGGTGCAGTGCGC	1457	
QY	796	CCAGCTCGAAAGTGTGATCGAGGCCACAGCTGCACGCGCAGCTGTTTGTGCAGCAGG	855	
DB	1458	CCAGCAAGAGGATCACCATTTCCCATGTAGCAGGTTCCACAGCTGGTGAATCGATGACG	1517	
QY	856	CAGGCCAGGCGGATGCCGGGAGTATAGCTGCGAGGCTGGGGCCA	901	
DB	1518	TCCGCCCGGAGGATGAGGAGACTACACGTTTGTGCTGACGGCTA	1563	

```

RESULT 8
US-09-902-540-7374/c
/ Sequence 7374, Application US/09902540
/ Patent No. 6833447
/ GENERAL INFORMATION:
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Wiegand, Roger C.
/ TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
/ FILE REFERENCE: 38-10(15849)B
/ CURRENT APPLICATION NUMBER: US/09/902,540
/ CURRENT FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: 60/217,883
/ PRIOR FILING DATE: 2000-07-10
/ NUMBER OF SEQ ID NOS: 16825
/ SEQ ID NO 7374
/ LENGTH: 690
/ TYPE: DNA
/ ORGANISM: Myxococcus xanthus
US-09-902-540-7374

```

Query Match	5.1%	Score 51.4	DB 4	Length 690
Best Local Similarity	44.1%	Pred. No. 0.034		
Matches 214	Conservative 0	Mismatches 271	Indels 0	Gaps 0
QY	279	CGCGGGTGCTCTTTGTGCGAGATGGCCCGGAGCATGCAGGCCCTCTACGAGTGGCTC	338	
DB	618	CAGCTTGTTCGCGAGCGCATGACGTGGGTGTCGATGGTCGGCTCTCCAGGAGGATGA	559	
QY	339	AGCCGCGGGGCGCATCGCTACCAGCTCTCCGTGCAAGGCCCTCGCGGCTTTCTGCAC	398	
DB	558	CAGGCCCCACACTCTTCACGAACTGCTCACGTGCTGCACCCGCGCCAGGCGGGCGAG	499	
QY	399	AAGACATGGCGGCGAGCTGTGTGAGATGCCGTGGCTGGGGGCCCGGCGCAGTTTGAGTGT	458	
DB	498	CAGGTCTCCAGAGCGCGAACTCCACGCGGTGAGGGGCCACCTCGCGGTCTCTCCACGAA	439	
QY	459	GAGACCTCCGAGGCCACGCTCCAGCTGCACTGGTACAAGGATGGCATGGAGCTGGGGCCAC	518	
DB	438	GAAACGGTGGGAGCGCCACGCTCAGGCGCAGGGGCCCCAGCGCCAGCGGCGCCCTGTC	379	
QY	519	TCGGGTGAGCGCTTCTTTGAGGAGGATGTGGGACCGCGCACCGGCTGGTGGCAGGCCACA	578	
DB	378	AGCGTGGGGCGCCCGCGCGGAGATGGCTTTCAGCGCAGCACCGACTCCCGCAGCT	319	
QY	579	GTCACGAGCAGGATGAAGGCACCTACTCTCTCGCGGTGGGCGAGACTCTGTGAGCTTC	638	
DB	318	GAAGGGCTTACACACGTAGTGTCTCGCGGCCACCTCGAAGCCCGGATGCGGTGCGACTC	259	
QY	639	CGGCTCGGCTCTCTAGAGCCCAAGTGGTGTGGCTTAAGGAGCAGCTGGGCACCCAGAAG	698	
DB	258	CTCGCCCTTGGCGGTGAGCATGACGATGAGGACGCCCCCGGGAATGGGGGTGTCGCCCGCAG	199	
QY	699	CTGACGCGAGAGCGGACGACGTGCCACACTCAGCTGCAGCTGGAGCTGGGCCACAGGCCAGC	758	
DB	198	CTGCCGCGACATCATGATGCCGACATGTCCGCGACATCAGTCCAGCAGCAGGAGTTC	139	
QY	759	GAGGT 763		
DB	138	GGGTT 134		

RESULT 9

```

US-09-902-540-1935/c
; Sequence 1935, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Goldmann, Gregory J.
; APPLICANT: Hinkle, Steven C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1935
; LENGTH: 2855
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1935

```

Query Match	5.1%	Score 51.4;	DB 4;	Length 2855;
Best Local Similarity	44.1%;	Pred. No. 0.045;		
Matches 214;	Conservative	0;	Mismatches 271;	Indels 0; Gaps 0;
QY	279	CGCGGGTGCTCTTTGCGAGATGTGGCCGGACGATCAGCGCTCTTACGAGTGCCTC	338	
Db	2804	CAGCTTTGTGCGCAGCGCATGACGTGGGTGTCGATGTGCGCTCTCCAGGGAGGATGA	2745	
QY	339	AGCGCGGGGGCGCGCATCGCCTACCAAGCTCTCGTGCAGAGCCTCGCGCGCTTTCTGCAC	398	

Db 2744 CAGGCCCCACACCTCTTCCAGCAACTGCTCACGTGTCTGACCCCGGCCAGCGGCGGAG 2685
QY 399 AAGGACATGGCGGCGAGCTGTGTGATGCGGTGGCTGGGGGCCCGCGCAGTTGAGTGT 458
Db 2684 CAGGTGCTCAGGAGCGCGAACTCCAGCGCGGTGAGGGCCACTCGCGGTCTCTCCACGAA 2625
QY 459 GAGACCTCCGAAGCCACAGCTCCACGTGCACTGTGTACAAGATGGCATGGAGCTGGGCCAC 518
Db 2624 GAAGCGGTGGAGCGCCACGCTCCAGGCGCAGGGGCCCCACGCCAGCGGCGCGTGCCTCC 2565
QY 519 TCGGTTGAGCGCTTCTTGACAGAGATGTGGGAACCGGCAACCGGTGTGTGACGCCACA 578
Db 2564 AGCCGTGGCGCGCGCGCGAGGATGCTTCAAGCCGACGACACAGCTCCCGCACGCT 2505
QY 579 GTCAACGAGCAGGATGAAGGCACCTACTCTCGCGGTGGCGGAGGACTCTGTGACTTC 638
Db 2504 GAAGGCTTACACGATGTCTCGCGGCCACCTCGAAGCCCGGATGCGGTTCGACTC 2445
QY 639 CGGCTCCGCTCTGTAGCCCCAAGTGTGTTTGTCTAAGGAGCAGCTGCGACGACGGAAG 698
Db 2444 CTCGCCCTTGGCGGTGAGCATGACGATGAGGACGCCCGGGACTGGGGCTGGCGCGAG 2385
QY 699 CTGCGAGGAGCGGAGGACCTGCTCCACACTGAGCTGGAGTGGCCAGGCCAGCAG 758
Db 2384 CTGCGGCGAGCATGATCGCGGACATGTCCGGCAGCATGTCCAGTCCAGCAGCAGGTC 2325
QY 759 GAGGT 763
Db 2324 GGGGT 2320

RESULT 10

US-09-902-540-1057
; Sequence 1057, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1057
; LENGTH: 10916
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1057

Query Match 5.1%; Score 51.4; DB 4; Length 10916;
Best Local Similarity 44.1%; Pred. No. 0.059;
Matches 214; Conservative 0; Mismatches 271; Indels 0; Gaps 0;
QY 279 CGCGGGTGTCTTGTGCGAGATGTGCGCGGACGATGCGCCCTCTACGAGTGGTC 338
Db 3371 CAGCTTGTGCGCGAGCGCATGACGTGGTGTGATGCTCTCCAGGAGGATGA 3430
QY 339 AGCGCGGGGCGCATGCTTACAGCTCTCCGTGCAAGCCCTCGCGCGCTTCTGAC 398
Db 3431 CAGCGCCACACCTCTTCCAGCAACTGCTCAGTGTCTGACCCGCGCGGCGGCGAG 3490
QY 399 AAGGACATGGCGGCGAGCTGTGTGATGCGCTGGGGCGCGCGAGTTTGTGTGT 458
Db 3491 CAGGTGTCCAGGAGCGGAACTCCAGCGGTGAGGGCCACCTCGCGGTCTCTCCAGAA 3550
QY 459 GAGACCTCCAGGCCACGCTCCAGTGTGCTGATGAGGATGGCATGGAGCTGGGCCAC 518
Db 3551 GAAGCGGTGGAGCCACGCTCCAGGCGAGGGGCCCCAGCGGCCAGCGCGCGCTGCCCC 3610

QY 519 TCGGTGAGCGCTTCTTTCAGGAGATGTGGGAGCGGCGACCGGTGGTGGCAGCCACA 578
Db 3611 AGCCGTGGCGCGCGCGGAGGATGGCTTCAGCGCGCAGACCACTCCCGACGCT 3670
QY 579 GTCAACGAGCAGGATGAAGGCACCTACTCTCGCGGTGGCGGAGGACTCTGTGACTTC 638
Db 3671 GAAGGCTTACACGAGTGTGCTCGCGGCCACCTCGAAGCCCGGATGCGGTGCGACTC 3730
QY 639 CGGCTCCGCTCTTGTGAGCCCCAAGTGTGTTTGTCTAAGGAGCAGCTTGGCAGCGGAAG 698
Db 3731 CTGCGCTTGGCGGTGAGCATGACGATGAGGACGCCCGCGACTGGGGGTGGCGCGAG 3790
QY 699 CTGCGAGCAGGAGGAGGAGCGGAGCCAGTCCACACTGAGCTGCGAGGTGGCGGCCAGCAG 758
Db 3791 CTGCGCGAGACATCGATGCGGACATGTCCGGCAGCATCAGGTCCAGCAGCAGGTC 3850
QY 759 GAGGT 763
Db 3851 GGGGT 3855

RESULT 11

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 5.1%; Score 51; DB 1; Length 7218;
Best Local Similarity 1.8%; Pred. No. 0.066;

; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match 4.9%; Score 49.4; DB 3; Length 50937;
Best Local Similarity 48.1%; Pred. No. 0.22;
Matches 140; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 76 GCATCCGAGTATGAGTCCGGGGCCCGCAGGGCTGACTGCCAACAGCCGCGCAGCCGAG 135
|||
Db 1146 GCAGACCATGAGGCGCTGTGAAACGCGGTGCCCATGTGCGGTGCGCAGATCCGGG 1205
|||
Qy 136 CTGCCGGGAGGTGCTGGCTCGCTGCACGAGGAGCGCAGCTGTGCTGAGCTGTCTAG 195
|||
Db 1206 AGCAGACATGAAACCCGAGCGATCGTCGAGCTGGGCTTCGGCCGGGCACATCCCGCGGG 1265
|||
Qy 196 ATCAGGCTCGGCTGTGACGTGGCTGAAAGGATGGTCGCACACTGTGCCCGAGGCCCGCAAGT 255
|||
Db 1266 ACCAGGTACGCGCCGAGAAAGCTGCGGAGGCGGTGCTCGCCGTGCGCTCCGACCCCGGTG 1325
|||
Qy 256 ATGAGGTGCGAGCATCGCGCGGGCGGGGTGCTCTCTGTGCGAGATGTGCGCCGCGGACG 315
|||
Db 1326 TCGCCGAACGGCTCGCGGCGCTTCGCGCAGAGATCCGTGAGCGCGGCGCGCGCGCGGG 1385
|||
Qy 316 ATCAGGCTCTTACAGTGGCTGACGCGCGGGGCGCATCGCTTACCAGC 366
|||
Db 1386 CGCCGACATCTGAGGGGATCTCGCCGAAGCAGGCTGACCGCCCTGC 1436
|||

RESULT 15
US-09-902-540-8525
; Sequence 8525, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8525
; LENGTH: 2670
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8525

Query Match 4.9%; Score 49; DB 4; Length 2670;
Best Local Similarity 43.8%; Pred. No. 0.15;
Matches 312; Conservative 0; Mismatches 395; Indels 6; Gaps 2;

Qy 85 TGGAGGTCCGGGGCGGCCCGGCTGACTGCCAACAGCCGACGCGAGCTGCCCGGG 144
|||
Db 1163 TGCTGGGCGCGCCCTGGAGCGCTCGAAGGACGCGACGCGGGACTTCGCGAGCTGCGGC 1222
|||
Qy 145 AGGTGCTGGCTCGCTGCAAGGAGGCGCAGCTGTGGCTGAGCTGTGAGATCAGGCTG 204
|||
Db 1223 AGGTCCGGGAGCGGGTGTCTGCGCTGGGACTGTCTGTTGGCGGAGCTGAGGTGCGCGCG 1282
|||
Qy 205 CGGCTGTGACGTGCTGAAGATGTCGCACACTGTCTCCAGGGCCCAAGTATGAGGTGC 264
|||
Db 1283 CGGCCGAGGACGCGGTGAGCGCGCGGCTCGTTCAACGGGGCGCCGCGCGCGAG 1342
|||
Qy 265 AGGCATCGCGCGCGCGGCTGTCTCTGTGCGAGATGTGGCCCGGAGCGATGACAGCC 324
|||

Db 1343 GTGGGGCGCGGCTCTCTGGAAGTGTGTCGCGCGGCTGAAGAGGCCAGTCGGAGTCGGTG 1402
Qy 325 TCTACGAGTGTGTCAGCGCGGGGGCGCATCGCTACAGCTCTCGTGTCAAGGCTCG 384
|||
Db 1403 AGTCCGCGTGGCGCACGCTCATCTCAGCATGGCCAGCAC--CGCCGAGGACGTGTCTG 1459
|||
Qy 385 CGGCTTTTCTGCAACAAGGACATGGCGGGCAGCTGTGTGGATGCGCTGGTGGGGGCCCGG 444
|||
Db 1460 CGGCTTTCCGCTGTTGAAGCACGCGGGCCCTGTGGGACGAGAGCGGGGCTGCGCCACGG 1519
|||
Qy 445 CGCAGTTTGAAGTGTGAGACCTCCGAAGCCCAAGCTCC--ACGTGCACTGTGTACAAGGATG 501
|||
Db 1520 CGAATGTGTGTCGCGCTGTTGAGCAGTTGGGCGCGCTGGGACTCGGGGCCCGACGTGCTGC 1579
|||
Qy 502 GCATGAGAGTGGGCGCACTCCGCTGAGCGCTTCTTGGCAGGAGGATGTGGGAGCGCGGACCC 561
|||
Db 1580 GCACGCTTTCGCGGACCGCGAGTACCGCAAGCACCTGGACGTGCGCGGTGGCCAGGAAG 1639
|||
Qy 562 GGCTGTGGCAGCCACAGTCCAGGCGAGGATGAAGGCACCTACTCTGCGCGCTGGGCG 621
|||
Db 1640 TGATGTGGCTACAGCGACTCCGGCAAGGAGGTGGGGCTCTTGGCGGCCAGCGCGCGC 1699
|||
Qy 622 AGGACTCTGTGGACTTCCGGCTCGCGCTCTCTGAGCCCAAGGTGTGTTTGTCTAAGAGC 681
|||
Db 1700 TGTACCGCGCGCAGGTGGCGCTCACGAGGTGTCTCGTGAAGCGCGGCGTCCCGCTGCGAC 1759
|||
Qy 682 AGCTGGCAGCGCAGGAAGCTGCAGGCGAGGCGAGGCGCAGTGCACACTGAGCTGCGAGG 741
|||
Db 1760 TCTTCCAGCGCCGCGGGAGTCCGTGGCGGTGCGGTGTCGCGCGCAGAGGCGGATTC 1819
|||
Qy 742 TGCGCCAGGCGCCAGACGAGGAGTGCAGTGTGTAACAAGGATGGGAAGAAGCTGAGC 794
|||
Db 1820 TCGCGCTGCGCGCGCGCGGTGGCGGTGGCTTACAAGGCGCGAGCAGGACAGGCG 1872
|||

Search completed: March 21, 2005, 14:39:03
Job time : 188.678 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2005, 17:32:32 ; Search time 627.052 Seconds
(without alignments)
9503.485 Million cell updates/sec

Title: US-10-077-130-4_COPY_10500_11500

Perfect score: 1001

Sequence: 1 gcccctgtgagtgaggaa.....ttctatagaagatgtgaaaaa 1001

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5544816 seqs, 2976611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1001	100.0	20489	18	US-10-723-860-5701 Sequence 5701, Ap
2	1001	100.0	23907	13	US-10-077-130-6 Sequence 6, Appli
3	1001	100.0	24120	13	US-10-077-130-4 Sequence 4, Appli
4	997.8	99.7	7564	18	US-10-476-397-15 Sequence 15, Appl
5	997.8	99.7	14061	17	US-10-093-463-73 Sequence 73, Appl
6	997.8	99.7	14109	17	US-10-093-463-71 Sequence 71, Appl
7	107.4	10.7	2534	17	US-10-108-260A-1946 Sequence 1946, Ap
8	62	6.2	3935	17	US-10-108-260A-534 Sequence 534, App
9	60	6.0	60	10	US-09-908-975-13516 Sequence 13516, A
10	59	5.9	82027	19	US-10-278-698-1034 Sequence 1034, Ap
11	59	5.9	82027	19	US-10-278-698-1045 Sequence 1045, Ap

12	57.4	5.7	81940	9	US-09-759-508B-1	Sequence 1, Appli
13	57.4	5.7	81940	10	US-09-960-706-1092	Sequence 1092, Ap
14	57.4	5.7	81940	10	US-09-873-319-724	Sequence 724, App
15	57.4	5.7	81940	18	US-10-723-860-132	Sequence 132, App
16	57.4	5.7	81940	18	US-10-723-860-132	Sequence 1, Appli
17	52.6	5.4	2170	17	US-10-264-049-850	Sequence 850, App
18	52.6	5.3	1115	14	US-10-184-644-440	Sequence 440, App
19	52.6	5.3	1115	14	US-10-184-644-440	Sequence 440, App
20	52	5.2	1645	9	US-09-726-643-22	Sequence 22, Appl
21	52	5.2	1645	13	US-10-042-141-22	Sequence 22, Appl
22	52	5.2	1645	19	US-10-919-272-22	Sequence 22, Appl
23	51	5.1	2335	17	US-10-424-599-81504	Sequence 81504, A
24	49.8	5.0	2254	9	US-09-726-643-44	Sequence 44, Appl
25	49.8	5.0	2254	13	US-10-042-141-44	Sequence 44, Appl
26	49.8	5.0	2254	19	US-10-919-272-44	Sequence 44, Appl
27	47.6	4.8	716	10	US-09-822-846-117	Sequence 117, App
28	46.8	4.7	671	14	US-10-184-644-346	Sequence 346, App
29	46.8	4.7	671	14	US-10-184-644-346	Sequence 346, App
30	45.6	4.6	594	14	US-10-123-155-10	Sequence 10, Appl
31	45.6	4.6	594	15	US-10-146-731-10	Sequence 10, Appl
32	45.6	4.6	594	15	US-10-140-472-10	Sequence 10, Appl
33	45.6	4.6	594	15	US-10-141-761-10	Sequence 10, Appl
34	45.6	4.6	594	16	US-10-142-885-10	Sequence 10, Appl
35	45.6	4.6	594	16	US-10-158-790-10	Sequence 10, Appl
36	45.6	4.6	594	17	US-10-137-871-10	Sequence 10, Appl
37	45.6	4.6	594	17	US-10-140-923-10	Sequence 10, Appl
38	45.6	4.6	594	17	US-10-141-756-10	Sequence 10, Appl
39	45.6	4.6	594	17	US-10-141-759-10	Sequence 10, Appl
40	45.6	4.6	594	17	US-10-140-805-10	Sequence 10, Appl
41	45.6	4.6	594	17	US-10-140-864-10	Sequence 10, Appl
42	45.6	4.6	594	17	US-10-142-426-10	Sequence 10, Appl
43	44	4.4	1257	18	US-10-472-380-4	Sequence 1, Appli
44	44	4.4	1435	18	US-10-472-380-4	Sequence 4, Appli
45	43.8	4.4	882	14	US-10-184-644-574	Sequence 574, App

ALIGNMENTS

```
RESULT 1
US-10-723-860-5701
; Sequence 5701, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nataasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723.860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 5701
; LENGTH: 20489
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (565)..(584)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2938)..(2955)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3329)..(3364)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
```

NAME/KEY: misc_feature
LOCATION: (4556)..(4586)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4721)..(4735)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4997)..(5011)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5108)..(5138)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (6237)..(6300)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (7520)..(7566)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (12014)..(12037)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (12123)..(12146)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (13258)..(13277)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (13918)..(13942)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15219)..(15248)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15262)..(15280)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15309)..(15381)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15862)..(15882)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (17286)..(17303)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (17413)..(17430)
OTHER INFORMATION: n is a, c, g, or t

Query Match 100.0%; Score 1001; DB 18; Length 20489;

Best Local Similarity 100.0%; Pred. No. 2.4e-268;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCTGTGGAGTGGAGAGGGGCCCGAGAACTCAGAGATGGGGACAGATACATCCTG 60
Db 10500 GCCCTGTGGAGTGGAGAGGGGCCCGAGAACTCAGAGATGGGGACAGATACATCCTG 10559

RESULT 2

US-10-077-130-6

; Sequence 6, Application US/10077130
; Publication No. US20020168742A1

; GENERAL INFORMATION:

; APPLICANT: Kapeller-Libermann, Rosana

; APPLICANT: Acton, Susan L.

; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family

Qy 61 AGGCAGAGGGGACCCAGTGTGAGCTGCAGATCTGTGGCTTGGCCATGGCGGACGCCGGG 120
Db 10560 AGGCAGAGGGGACCCAGTGTGAGCTGCAGATCTGTGGCTTGGCCATGGCGGACGCCGGG 10619
Qy 121 GAGTACTTGTGTGTGGGGCAGGAGAGACCTCAGCCAGCTCACCATCAGGGCTCTG 180
Db 10620 GAGTACTTGTGTGTGGGGCAGGAGAGACCTCAGCCAGCTCACCATCAGGGCTCTG 10679
Qy 181 CCTGCCAGGTTTCATAGAAGATGTGAAAAACCCAGAGGCCAGAGAGGGGCCACCGCTGTG 240
Db 10680 CCTGCCAGGTTTCATAGAAGATGTGAAAAACCCAGAGGCCAGAGAGGGGCCACCGCTGTG 10739
Qy 241 CTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
Db 10740 CTGCAGTGTGAGCTGAACAGTGCAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10799
Qy 301 AGAGATGGGACAGATACAGCTTGGCAGGACGGGACTAAATGTGTGTGTGTGTGTGTGTGT 360
Db 10800 AGAGATGGGACAGATACAGCTTGGCAGGACGGGACTAAATGTGTGTGTGTGTGTGTGTGT 10859
Qy 361 GGCTTGGCCATGGCAGACACTGGGGAGTACTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
Db 10860 GGCTTGGCCATGGCAGACACTGGGGAGTACTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10919
Qy 421 GCTATGCTCAACCGTCAAGGCTCTACCCATCAAGTTTCACAGAGGGTCTGTGTGTGTGTGTGT 480
Db 10920 GCTATGCTCAACCGTCAAGGCTCTACCCATCAAGTTTCACAGAGGGTCTGTGTGTGTGTGTGT 10979
Qy 481 GCCACAGAGGGGCAACAGCCGT 540
Db 10980 GCCACAGAGGGGCAACAGCCGT 11039
Qy 541 TGGTGAAGGGGCATGAGACCTCAGAGATGGAGACAGACACAGCCCTCAGGCGAGGACGGG 600
Db 11040 TGGTGAAGGGGCATGAGACCTCAGAGATGGAGACAGACACAGCCCTCAGGCGAGGACGGG 11099
Qy 601 GCCAGTGTGAGCTGCAGATCCCGGCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
Db 11100 GCCAGTGTGAGCTGCAGATCCCGGCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11159
Qy 661 ATGTGGGGAAGGAGGAGGACCTCAGCCATGCTCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
Db 11160 ATGTGGGGAAGGAGGAGGACCTCAGCCATGCTCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11219
Qy 721 ATAGAGGCTCTGAGGAATGAAGAGGCCACAGAGGGGACACGCGACGCTGTGTGTGTGTGTGTGTGT 780
Db 11220 ATAGAGGCTCTGAGGAATGAAGAGGCCACAGAGGGGACACGCGACGCTGTGTGTGTGTGTGTGTGT 11279
Qy 781 CTGAGCAAGCGGCAACCGGTGGAGTGGAGAGGGGCAAGAGGAGGACCTCAGAGATGGGGAC 840
Db 11280 CTGAGCAAGCGGCAACCGGTGGAGTGGAGAGGGGCAAGAGGAGGACCTCAGAGATGGGGAC 11339
Qy 841 AGACACAGCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
Db 11340 AGACACAGCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11399
Qy 901 GTGGATGCCGGGAGTACTCGT 960
Db 11400 GTGGATGCCGGGAGTACTCGT 11459
Qy 961 GTCAGGGCCCTGCTGCCAGATTCATAGAAGATGTGAAAAA 1001
Db 11460 GTCAGGGCCCTGCTGCCAGATTCATAGAAGATGTGAAAAA 11500

;; TITLE OF INVENTION: Members and Uses Therefor
;; FILE REFERENCE: MPI2001-047P1RCP1(M)
;; CURRENT APPLICATION NUMBER: US/10/077,130
;; CURRENT FILING DATE: 2002-02-15
;; PRIOR APPLICATION NUMBER: 60/269201
;; PRIOR FILING DATE: 2001-02-15
;; NUMBER OF SEQ ID NOS: 9
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6
;; LENGTH: 23907
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match 100.0%; Score 1001; DB 13; Length 23907;
Best Local Similarity 100.0%; Pred. No. 2.4e-268;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCCTGTGGAGTGGAGGAGGCGCCGAGAACCTCAGAGATGGGACAGATACATCCTG 60
Db 10429 GCCCCTGTGGAGTGGAGGAGGCGCCGAGAACCTCAGAGATGGGACAGATACATCCTG 10488

Qy 61 AGGCAGGAGGGACACAGGTGTGAGCTGCAGATCTGTGGCCTGGCCATGGCGGACGCCGGG 120
Db 10489 AGGCAGGAGGGACACAGGTGTGAGCTGCAGATCTGTGGCCTGGCCATGGCGGACGCCGGG 10548

Qy 121 GAGTACTTGTGTGTGCGGCGAGAGAGGACCTCAGCCACGCTCACCATCAGGGCTCTG 180
Db 10549 GAGTACTTGTGTGTGCGGCGAGAGAGGACCTCAGCCACGCTCACCATCAGGGCTCTG 10608

Qy 181 CCTCCAGGTTTCATAGAGATGTGAAACACAGAGGCGCAGAGAGGGCCACCGCTGTG 240
Db 10609 CCTCCAGGTTTCATAGAGATGTGAAACACAGAGGCGCAGAGAGGGCCACCGCTGTG 10668

Qy 241 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGAGTGGAGAGAGGGGTCTGAGACCCCTC 300
Db 10669 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGAGTGGAGAGAGGGGTCTGAGACCCCTC 10728

Qy 301 AGAGATGGGACAGATACAGCTTGAGGCGAGGAGCGGACCTAAATGTAGCTGCAGATTCTG 360
Db 10729 AGAGATGGGACAGATACAGCTTGAGGCGAGGAGCGGACCTAAATGTAGCTGCAGATTCTG 10788

Qy 361 GGCCTGGCCATGGCAGACACTGGGAGTACTCTGCTGCGTGTGCGGCGAGGAGGACCTCG 420
Db 10789 GGCCTGGCCATGGCAGACACTGGGAGTACTCTGCTGCGTGTGCGGCGAGGAGGACCTCG 10848

Qy 421 GCTATGCTCACCGTCAAGGCTCTACCCATCAAGTTTCAGAGAGGGTCTCAGGAAACGAGAG 480
Db 10849 GCTATGCTCACCGTCAAGGCTCTACCCATCAAGTTTCAGAGAGGGTCTCAGGAAACGAGAG 10908

Qy 481 GCCACAGAGGGGCAACAGCCGTGTGCGGTGTGAGCTGAGCAAGATGGCCCGCTGGAG 540
Db 10909 GCCACAGAGGGGCAACAGCCGTGTGCGGTGTGAGCTGAGCAAGATGGCCCGCTGGAG 10968

Qy 541 TGGTGAAGGGGCGATGACACCTCAGAGATGGAGACAGACACAGCCTGAGCAGGACGGG 600
Db 10969 TGGTGAAGGGGCGATGACACCTCAGAGATGGAGACAGACACAGCCTGAGCAGGACGGG 11028

Qy 601 GCCAGGTGTGAGCTGCAGATCCGGGCTCTGCGGCGAGAGGACGCTGGGGAGTACTCTGTC 660
Db 11029 GCCAGGTGTGAGCTGCAGATCCGGGCTCTGCGGCGAGAGGACGCTGGGGAGTACTCTGTC 11088

Qy 661 ATGTGCGGGAAGGAGGAGGACCTCAGCCATGTCTACCGTCAAGGGCCATCGCTTCAAGTTC 720
Db 11089 ATGTGCGGGAAGGAGGAGGACCTCAGCCATGTCTACCGTCAAGGGCCATCGCTTCAAGTTC 11148

Qy 721 ATAGAGGTCTGAGGAATGAAGGCCACAGAGGGGACAGGCCACGCTGTGTTGAG 780
Db 11149 ATAGAGGTCTGAGGAATGAAGGGCCACAGAGGGGACAGGCCACGCTGTGTTGAG 11208

Qy 781 CTGAGCAAGCGGCGACCGGTGAGTGGAGGAGGGGCATGAGCCTCAGAGATGGGGAC 840
Db 11209 CTGAGCAAGCGGCGACCGGTGAGTGGAGGAGGGGCATGAGCCTCAGAGATGGGGAC 11268

Qy 841 AGACACAGCCTGAGGACAGGACGGGTCCAGGTGTGAGCTGCAGATCCGTGGCCTGGCTGTG 900
Db 11269 AGACACAGCCTGAGGACAGGACGGGTCCAGGTGTGAGCTGCAGATCCGTGGCCTGGCTGTG 11328

Qy 901 GTGGATCCCGGGGAGTACTCTGTGTGTGCGGGCAGGAGAGGACCTCAGCCACACTCACT 960
Db 11329 GTGGATCCCGGGGAGTACTCTGTGTGTGCGGGCAGGAGAGGACCTCAGCCACACTCACT 11388

Qy 961 GTCAGGCGCCTGCTGCCAGATTCATAGAAAGATGTGAAAA 1001
Db 11389 GTCAGGCGCCTGCTGCCAGATTCATAGAAAGATGTGAAAA 11429

RESULT 3
US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-4

Query Match 100.0%; Score 1001; DB 13; Length 24120;
Best Local Similarity 100.0%; Pred. No. 2.4e-268;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCCTGTGGAGTGGAGGAGGCGCCGAGAACCTCAGAGATGGGACAGATACATCCTG 60
Db 10500 GCCCCTGTGGAGTGGAGGAGGCGCCGAGAACCTCAGAGATGGGACAGATACATCCTG 10559

Qy 61 AGGCAGGAGGGACACAGGTGTGAGCTGCAGATCTGTGGCCTGGCCATGGCGGACGCCGGG 120
Db 10560 AGGCAGGAGGGACACAGGTGTGAGCTGCAGATCTGTGGCCTGGCCATGGCGGACGCCGGG 10619

Qy 121 GAGTACTTGTGTGTGCGGCGAGGAGGACCTCAGCCACGCTCACCATCAGGGCTCTG 180
Db 10620 GAGTACTTGTGTGTGCGGCGAGGAGGACCTCAGCCACGCTCACCATCAGGGCTCTG 10679

Qy 181 CCTGCCAGGTTTCATAGAGATGTGAAAAACAGGAGGCGCAGAGAGGGCCACCGCTGTG 240
Db 10680 CCTGCCAGGTTTCATAGAGATGTGAAAAACAGGAGGCGCAGAGAGGGCCACCGCTGTG 10739

Qy 241 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGAGAGTGGAGAAAGGGGTCTGAGACCCCTC 300
Db 10740 CTGCAGTGTGAGCTGAACAGTGCAGCCCTCTGAGAGTGGAGAAAGGGGTCTGAGACCCCTC 10799

Qy 301 AGAGATGGGACAGATACAGCCTGAGGCGAGGAGGACCTCAGCCACGCTCAATGTAGCTGCAGATTCTG 360
Db 10800 AGAGATGGGACAGATACAGCCTGAGGCGAGGAGGACCTCAGCCACGCTCAATGTAGCTGCAGATTCTG 10859

Qy 361 GGCCTGGCCATGGCAGACACTGGGAGTACTCTGCTGCGTGTGCGGCGAGGAGGACCTCG 420

10860	Db	GGCCTGGCCATGCACAGACACTGGGGAGTACTCGTGGTGTGCGGCGAGAGAGCACTCG	10919
421	Qy	GCTATGCTCACCGTCAGGGCTCTACCCATCAAGTTTCACAGAGGGTCTGAGGAACGAAG	480
10920	Db	GCTATGCTCACCGTCAGGGCTCTACCCATCAAGTTTCACAGAGGGTCTGAGGAACGAAG	10979
481	Qy	GCACAGAGGGGCAACAGCGGTGTCGGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG	540
10980	Db	GCACAGAGGGGCAACAGCGGTGTCGGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG	11039
541	Qy	TGCTGGAAGGGGCATCAGACCCCTCAGAGATGGAGACAGACACAGCTCAGGGCAGCACGG	600
11040	Db	TGCTGGAAGGGGCATCAGACCCCTCAGAGATGGAGACAGACACAGCTCAGGGCAGCACGG	11099
601	Qy	GCCAGGTGTGAGCTGCAGATCCGCGGCTCTCGTGGCAGAGGACGCTGGGAGTACCTGTGC	660
11100	Db	GCCAGGTGTGAGCTGCAGATCCGCGGCTCTCGTGGCAGAGGACGCTGGGAGTACCTGTGC	11159
661	Qy	ATGTGCGGGAAGAGAGGACCTCAGCCATGCTCACCGTCAGGGCCATGCTTCCAAGTTC	720
11160	Db	ATGTGCGGGAAGAGAGGACCTCAGCCATGCTCACCGTCAGGGCCATGCTTCCAAGTTC	11219
721	Qy	ATAGAGGGTCTGAGGAATGAAGAGGCCACAGAAGGGGACACGGCCACGCTGTGGTGCAG	780
11220	Db	ATAGAGGGTCTGAGGAATGAAGAGGCCACAGAAGGGGACACGGCCACGCTGTGGTGCAG	11279
781	Qy	CTGAGCAAGCGGCACCGGTGGAGTGAGGAAGGGGCATGAGACCCCTCAGAGATGGGAC	840
11280	Db	CTGAGCAAGCGGCACCGGTGGAGTGAGGAAGGGGCATGAGACCCCTCAGAGATGGGAC	11339
841	Qy	AGACACAGCCTGAGGCAGACCGGCTCAGGTGTGAGCTGCAGATCCGTGGCCCTGGCTGTG	900
11340	Db	AGACACAGCCTGAGGCAGACCGGCTCAGGTGTGAGCTGCAGATCCGTGGCCCTGGCTGTG	11399
901	Qy	GTGGATCCGGGGAGTACTCGTGTGTGTCGGGGCAGGAGGACCTCAGGCCACTCACT	960
11400	Db	GTGGATCCGGGGAGTACTCGTGTGTGTCGGGGCAGGAGGACCTCAGGCCACTCACT	11459
961	Qy	GTGAGGGCCCTGCTGCCAGATTCATAGAAGATGTGAAAA	1001
11460	Db	GTGAGGGCCCTGCTGCCAGATTCATAGAAGATGTGAAAA	11500

RESULT 4

US-10-476-397-15
; Sequence 15, Application US/10476397
; Publication No. US20040115687A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: LEE, Ernestine A.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: THANGAVELOU, Kavitha
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: DING, Li
; APPLICANT: JACKSON, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: LEE, Sally
; APPLICANT: WARREN, Bridget A.
; APPLICANT: XU, Yuming
; APPLICANT: TRAN, Uyen K.
; APPLICANT: LAL, Preeti G.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: HAFALIHA, April J.A.
; APPLICANT: YAO, Monique G.
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: CHINN, Anna M.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: RAMKUMAR, Jayalaxmi

Db 3021 GGCCTGGCCATGGCAGACACTGGGGAGTACTCGTGGTGTGCGGCACGAGAGGACCTCG 3080
QY 421 GCTATGCTCACCGTCAAGGCTCTACCCATCAAGTTTCACAGAGGGTCTGAGGAACGAAGAG 480
Db 3081 GCTATGCTCACCGTCAAGGCTCTACCCATCAAGTTTCACAGAGGGTCTGAGGACGAAGAG 3140
QY 481 GCCACAGAAGGGGCAACAGCCGTGTGCGGTGTGAGCTGTGAGCAAGATGGCCCCCGTGGAG 540
Db 3141 GCCACAGAAGGGGCAACAGCCGTGTGCGGTGTGAGCTGTGAGCAAGATGGCCCCCGTGGAG 3200
QY 541 TGGTGGAGGGGCATGAGACCTCTCAGAGATGAGAGACAGACACAGCCTGAGGCAGACGGG 600
Db 3201 TGGTGGAGGGGCATGAGACCTCTCAGAGATGAGAGACAGACACAGCCTGAGGCAGACGGG 3260
QY 601 GCCAGGTGTGAGCTGTGAGATCCGCGCCCTCGTGGCAGAGGACGCTGGGGAGTACCTGTGC 660
Db 3261 GCCAGGTGTGAGCTGTGAGATCCGCGCCCTCGTGGCAGAGGACGCTGGGGAGTACCTGTGC 3320
QY 661 ATGTGCGGGAAGGAGGAGACCTTCAGCCATGCTTCACCGTCAGGGCCATGCCCTTCCAAGTTC 720
Db 3321 ATGTGCGGGAAGGAGGAGACCTTCAGCCATGCTTCACCGTCAGGGCCATGCCCTTCCAAGTTC 3380
QY 721 ATAGAGGGTCTGAGGAATGAGAGGCCACAGAGGGGACACGGCCACGCTGTGGTGTGAG 780
Db 3381 ATAGAGGGTCTGAGGAATGAGAGGCCACAGAGGGGACACGGCCACGCTGTGGTGTGAG 3440
QY 781 CTGAGCAAGCGGCGCACCGGTGTGAGTGGAGGAGGAGGCGCATGAGACCCCTCAGAGATGGGGAC 840
Db 3441 CTGAGCAAGCGGCGCACCGGTGTGAGTGGAGGAGGAGGCGCATGAGACCCCTCAGAGATGGGGAC 3500
QY 841 AGACACAGCCTGAGCAGACGAGGCTCAGGTTCCAGGTGTGAGCTGTGAGATCCGTGGGCTGGCTGTG 900
Db 3501 AGACACAGCCTGAGCAGACGAGGTTCCAGGTGTGAGCTGTGAGATCCGTGGGCTGGCTGTG 3560
QY 901 GTGGATGCGGGGAGTACTCGTGTGTGTGCGGAGGAGAGGACCTCAGGCCACACTCACT 960
Db 3561 GTGGATGCGGGGAGTACTCGTGTGTGTGCGGAGGAGAGGACCTCAGGCCACACTCACT 3620
QY 961 CTCAGGGCCCTGCCCTGCCAGATTCATAGAAGATGTGAAAA 1001
Db 3621 CTCAGGGCCCTGCCCTGCCAGATTCATAGAAGATGTGAAAA 3661

RESULT 6

US-10-093-463-71
; Sequence 71, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Raastelli, Luca
; APPLICANT: Meres, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zernhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward

Query Match 99.7%; Score 997.8; DB 17; Length 14109;
Best Local Similarity 99.8%; Pred. No. 1.8e-267;
Matches 999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCCCTGTGTGAGTGGAGGAGGGCCCGAGAACCTCAGAGATGGGACAGATACATCCTG 60
Db 2661 GCCCTGTGTGAGTGGAGGAGGGCCCGAGAACCTCAGAGATGGGACAGATACATCCTG 2720
QY 61 AGCAGAGGGGACACAGGTGTGAGCTGCAGATCTGTGGCTTGGCCATGGGACGCGGG 120
Db 2721 AGCAGAGGGGACACAGGTGTGAGCTGCAGATCTGTGGCTTGGCCATGGGACGCGGG 2780
QY 121 GAGTACTTGTGTGTGCGGGCAGGAGAGGACCTCAGCCACGCTCACCATCAGGGGCTGTG 180
Db 2781 GAGTACTTGTGTGTGCGGGCAGGAGAGGACCTCAGCCACGCTCACCATCAGGGGCTGTG 2840

181 CCTGCCAGGTTTCATAGAGATGTGAAAAACAGAGAGCCAGAGAGGGCCACCGCTGTG 240
Db |||||
2841 CCTGCCAGGTTTCATAGAGATGTGAAAAACAGAGAGCCAGAGAGGGCCACCGCTGTG 2900
Qy |||||
241 CTGCAAGTGTAGCTGAACAGTGCAGCCCTCTGGAGTGTGAGAAAGGGTCTGAGACCTC 300
Db |||||
2901 CTGCAAGTGTAGCTGAACAGTGCAGCCCTCTGGAGTGTGAGAAAGGGTCTGAGACCTT 2960
Qy |||||
301 AGAGATGGGGACAGATACAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
Db |||||
2961 AGAGATGGGGACAGATACAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3020
Qy |||||
361 GGCCTGGCCATGGGACAGACACTGGGAGTACTCGTGGTGTGGGAGGAGGAGGAGGAGGAG 420
Db |||||
3021 GGCCTGGCCATGGGAGACACTGGGAGTACTCGTGGTGTGGGAGGAGGAGGAGGAGGAG 3080
Qy |||||
421 GCTATGCTCACCGTCAAGGCTCTACCCATCAAGTTTACAGAGGGTCTGAGGAAACGAGAG 480
Db |||||
3081 GCTATGCTCACCGTCAAGGCTCTACCCATCAAGTTTACAGAGGGTCTGAGGAAACGAGAG 3140
Qy |||||
481 GCCACAGAGGGGCAACAGCGTGTCTGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG 540
Db |||||
3141 GCCACAGAGGGGCAACAGCGTGTCTGGTGTGAGCTGAGCAAGATGGCCCCCGTGGAG 3200
Qy |||||
541 TGGTGAAGGGGCAATGACACCTCAGAGATGGAGATGGAGACACACAGCCTCAGGACGAGCGG 600
Db |||||
3201 TGGTGAAGGGGCAATGACACCTCAGAGATGGAGATGGAGACACACAGCCTCAGGACGAGCGG 3260
Qy |||||
601 GCCAGGTGTAGCTGACATCGCGGCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
Db |||||
3261 GCCAGGTGTAGCTGACATCGCGGCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3320
Qy |||||
661 ATGTGCGGGAAGGAGAGACCTCAGCCATGTCTACCTCAGGAGGAGGAGGAGGAGGAGGAG 720
Db |||||
3321 ATGTGCGGGAAGGAGAGACCTCAGCCATGTCTACCTCAGGAGGAGGAGGAGGAGGAGGAG 3380
Qy |||||
721 ATAGAGGTTCTGAGGATTAAGAGCCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
Db |||||
3381 ATAGAGGTTCTGAGGATTAAGAGCCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3440
Qy |||||
781 CTGAGCAAGGGCGGACCGGTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Db |||||
3441 CTGAGCAAGGGCGGACCGGTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3500
Qy |||||
841 AGACACAGCCTGAGCAGGAGCGGTCCAGGTGTGAGTGTGAGCTGAGATCCGTTGGCTGCTGTG 900
Db |||||
3501 AGACACAGCCTGAGCAGGAGTGGTCCAGGTGTGAGTGTGAGCTGAGATCCGTTGGCTGCTGTG 3560
Qy |||||
901 GTGATGCGGGGAGTACTCGTGTGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
Db |||||
3561 GTGATGCGGGGAGTACTCGTGTGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3620
Qy |||||
961 GTGATGCGGGGAGTACTCGTGTGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1001
Db |||||
3621 GTGATGCGGGGAGTACTCGTGTGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3661

RESULT 7

US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-108-260A-1946

Query Match 10.7%; Score 107.4; DB 17; Length 2534;
Best Local Similarity 47.7%; Pred. No. 6.5e-20;
Matches 420; Conservative 0; Mismatches 446; Indels 15; Gaps 3;
Qy 136 TGGCGGAGGAGAGGAGCCTCAGCCAGCTCACCATCAGGGCTCTCCCTGCCAGGTTTCATA 195
Db |||||
199 TCCCTGAAGTGTGACCTCTGCCCGCTGGAGGTCGAGTGAGCCCGTGGTGTCTCTG 258
Qy |||||
196 GAAGATGTGAAAAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 254
Db |||||
259 AAGCGCTGGATGACCTGTCCGAGAGGAGCGCGCACCTTGGCCCTCGAGTGTGAAGTC 318
Qy |||||
255 --GAAACAGTGCAGCCCTGTGAGAGTGGAGAAAGGGGTCTGAGACCTCAGAGATGGGAGC 312
Db |||||
319 TCTGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 378
Qy |||||
313 AGATACAGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 372
Db |||||
379 AAGTATGACTTCTGACACACGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 438
Qy |||||
373 GCAGACACTGGGAGGAGTACTCGTGGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 432
Db |||||
439 GAAGACGCGGCTGTACACCTGCCAGTGGGCTCCGAGGAGACCCGCGGCTCCGCG 498
Qy |||||
433 GTCAGGCTCTPACCCATCAAGTTTACAGAGAGGTCTGAGGAAACGAGAGGAGGAGGAGGAGGAGGAG 492
Db |||||
499 GTGACGATCTGACGCTGGGATCACCAAGAGGCTGAAGACAAATGGAGGTGCTGGAAGGG 558
Qy |||||
493 GCAACAGCGTGTGCGGTGTGAGCTGAGCAAGATGGCCCCGCTGGAGTGGTGGAGGAGGAGGAGGAG 552
Db |||||
559 GAAAGCTGCAGCTTGTGAGTGGCTCTGTCACAGAGTGCACGAGGAGGAGGAGGAGGAGGAGGAGGAG 618
Qy |||||
553 CATGAGACCTTCAGAGATGGAGACAGACACAGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 603
Db |||||
619 ACAGTGGTGGAGAGACAGTGGGAGCTCCAGCGCTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 678
Qy |||||
604 AGGTGTGAGTGTGAGATCCGCGGCTCTGTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 663
Db |||||
679 AAATACATCTCTGGTGGTCCGCGAGGCTGCACCAAGTGTATGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 738
Qy |||||
664 TGGGAG 723
Db |||||
739 GTGCGGGGCTTCACCTCCAGGCTCTCATCTGTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 798
Qy |||||
724 GAGGCTCTGAGGAATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 783
Db |||||
799 AAGCCCTGGAGACACAGTGGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 858
Qy |||||
784 AGCAAGGCGG---CACCGGTGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Db |||||
859 TCAGGCGGAG 918
Qy |||||
841 AGACACAGCCTGAG 900
Db |||||
919 AAGTATGATGTGTCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 978
Qy |||||
901 GTGGATGCGGGAGGAGTACTCGTGTGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
Db |||||
979 AAGGACGCGGAGGAGTACAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1038
Qy |||||
961 GTGAGGCGCTGCTGCCAGATTTCATAGAGATGTGAAAAA 1001
Db |||||
1039 GTGGAAGAAAGCAAACTGCTTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1079

RESULT 8

US-10-108-260A-534
; Sequence 534, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: No. US200400055601el full length cDNA
 ; FILE REFERENCE: HI-A0106
 ; CURRENT APPLICATION NUMBER: US/10/108,260A
 ; CURRENT FILING DATE: 2002-03-27
 ; NUMBER OF SEQ ID NOS: 5458
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 534
 ; LENGTH: 3935
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-108-260A-534

Query Match 6.2%; Score 62; DB 17; Length 3935;
 Best Local Similarity 46.0%; Pred. No. 2.8e-07;
 Matches 295; Conservative 0; Mismatches 335; Indels 12; Gaps 2;

Qy	309	GGACAGATACAGCTGAGCAGGACGGGACTAAATGTGACGTGCAGATTCTGTGCCTGGC 368
Db	1033	GGGCAAGTACGATGTGAAGCAGATGGGCAACCAAGTACATGCTGTGTTATTAGCAACGTGAA 1092
Qy	369	CATTGGCAGACACTGGGGAGTACTCGTGGGTGCGGCGAGGAGGACCTCGGCTATGCT 428
Db	1093	CATGAACGATGCTGGCATCTACACCTGTCCGTGGCGCATGAACGATGAGTGACAGCT 1152
Qy	429	CACCGTCAGGGCTCTACCCATCAAGTTCAACAGGGTCTGAGGAAACGAAAGGCCACAGA 488
Db	1153	CACAGTCTGGATGAGCCACTGAAGTTCTTGGGAGAGATGAAGCTGTGAAGGTGACAGA 1212
Qy	489	AGGGCAACAGCCGTGCTGGGTGTGAGCTGAGCAAGATGCCCCCGTGGAGTGGTGAA 548
Db	1213	CGGCCACACAGCTGTGTTGAGATCCGCTCTCCAAAGAAAGAGCCCAACTTTTGTGTGAA 1272
Qy	549	GGGCGATGAGACCCCTCAGAGATGAGACAGACA-----CAGCCTGAGGCAGGACGG 599
Db	1273	GTTCATTTGGGAAGGAGCTGAAGGGATGACAAGATGAATAATCAGGTGTCCGAAGATGG 1332
Qy	600	GGCCAGGTGTGAGCTGCAGATCCCGGCTCGTGGCAGAGCAGCTGGGGAGTACCTGTG 659
Db	1333	TCTGACGCACACGCTTAAGATTAAAGATGCCAGACTCAGTGACAGCGCGAGTTCTCTGC 1392
Qy	660	CATGTGGGGAAGAGAGGACCTCAGGCATGCTCAGCGTCAGGGCCATGCCTTCAAGTT 719
Db	1393	TGAGGCGGGGAAACCTGGTGTACAAAAGGCCACGCTCACTATTGACCGCATCCCCATCAAGTT 1452
Qy	720	CATAGAGGTCTGAGGAATGAAGAGGCCACAGAAGGGGACACGCCACGCTGTGGTGTA 779
Db	1453	TGTGAGCAACTCAAAATGTACTGTGTGAAGAGAGGAGTCCGCATGCTCTGGAGTGTA 1512
Qy	780	GCTGAC---GCAAGCGGCACCGGTGGAGTGGAGGAAGGGGCATGAGACCTCAGAGATGG 836
Db	1513	GCTGACATCCAAGATGTGACACTCGCTGGAGAGGATGGGAGCTGTGATGCATGG 1572
Qy	837	GGACAGACACAGCTGAGGAGGACGGGTCCAGGTGTGAGCTGCAGATCCGTGGCCTGGC 896
Db	1573	CACCTAAGTACAGCATGAACCATGAGGGCAAGCAAGCAGAGCTGATCATCGAGGATGCACA 1632
Qy	897	TGTGGTGGATCCCGGGAGTACTCGTGTGCTGGCGGCAGGA 938
Db	1633	GCTCAGTGTGTTGGCGAGTACACTGTGGTGGCCATGCAGGA 1674

RESULT 9
 US-09-908-975-13516
 ; Sequence 13516, Application US/09908975
 ; Publication No. US20030165843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SHOSHAN, Avi
 ; APPLICANT: WASSERMAN, Alon
 ; APPLICANT: MINTZ, Eli
 ; APPLICANT: MINTZ, Liat
 ; APPLICANT: FAIGLER, Simchon
 ; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICING
 ; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME

[illegible]

RESULT 11

```

US-10-278-598-1045
; Sequence 1045, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmüller, Bruno
; APPLICANT: Haupl, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: Q30027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1045
; LENGTH: 82027
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-598-1045

```

Query Match	5.9%	Score 59	DB 19	Length 82027
Best Local Similarity	43.4%	Pred. No. 2.3e-06		
Matches 376	Conservative 0	Mismatches 485	Indels 6	Gaps 2
Qy	141	GCAGGAGAGACCTCAGGCACGCTCACCATCAGCGCTCGCTGCCAGGTTTCATAGAAGA	200	
Db	15676	GGAGCTTGAGTCCACAGCACAGCTGAAGTCTTTAGAGCCGATCCCTACTTCACTGTGAA	15735	
Qy	201	TGTGAAAAACGAGGAGCCAGAGAAGGGGCCACGGCTGTGCTGCAGTGTGAGCTGAACAG	260	
Db	15736	ATTACATGACAAACTGCAGTGGAGAAGGATGAGATTACTTTGAAGTGTGAAGTCAGCAA	15795	
Qy	261	TGCAGCCCTGTGGAGTGGANAAGGGTCTGAGACCCCTCAGAGATGGGACAGATACAG	320	
Db	15796	AGATGTACCAAGTGAATGGTTCAAGATGGTGAAGAGATTGTCCCTTCACCCCAATATTC	15855	
Qy	321	CCTGAGGCGAGGACGGGACTAAATGTGAGCTGCAGATTTCGTGGCCCTGGCCATGGCAGACAC	380	

Db	15856	TATCAAGGCGAGATGGCGCTCGCGCGCATTTTAAAAATCAAAAAGGCGGACCTTAAAGATAA	15915
Qy	381	TGGGGAGTACTCGTGGTGTGCGGGCAGGAGGACCTCGGCTATGCTCACCGTCAGGGC	440
Db	15916	AGCGAATATGTGTGACTGTGGCCACAGACAAGCAATGTTACTGTTGAGGC	15975
Qy	441	TCTACCCATCAAGTTTCAAGAGGGTCTGAGGAAACGAAGGCCACAGAAAGGGGCACAGC	500
Db	15976	TCGACTAAATAAAGTGGAAAAGCCCTGTGACGGAGTAGAGGTGTTGTGTGGTGAACACAGC	16035
Qy	501	CGTGTCTGCGGTGTGAGCTGAGCAAGATGSCCCCGCTGGAGTGGTGGAAAGGGGCATG---A	557
Db	16036	CACTTTGAAATTTGAACCTTCTGAACTGATGTTTCAGCGCCAGTGGAAAGCTTGAAGGACA	16095
Qy	558	GACCTCTCAGAGATGGAGACAGACACAGCCCTCAGGCAGGACGGGGCCAGGTGTGAGCTGCA	617
Db	16096	GCCTTTTGACAGCTTCCCTCGACTGTGTGAATCATTTGAGGATGGAAGAAGCATATTTCTGAT	16155
Qy	618	GATCCGGCGCTGTGGCAGAGAGACGCTGGGGAGTACCTGTGCATGTGCGGGAGGAGAG	677
Db	16156	CCTTCTAACTGTGCACTGAGCTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGTAATGCCAA	16215
Qy	678	GACCTCAGCCCATGCTCACCGTCAGGGCCATGCTTCCAAGTTTCATAGAGGCTCTGAGGAA	737
Db	16216	ATCTGACGCCAATCTGAAAGTGAAGAATTTGCCTCTTATCTTCATCACACCTCTCAGTGA	16275
Qy	738	TGAAGAGGCCACAGAAGGGGACACAGGCCACGCTGTGGTGTGAGCTTGACCAAGGGCGC---	794
Db	16276	TGTTAAAGTCTTCGAGAAAGATGAGGCTAAGTTTCAGTGTGNAGTATCCAGGAGGCCAA	16335
Qy	795	ACCGTGGAGTGGAGAAAGGGGCATGAGACCTTCAGAGATGGGACAGACACAGCCTGAG	854
Db	16336	AACATTTCCGTTGGCTAAAAGGAAACCCAGGAATACAGGTGATGACAGATTGAGCTTTAT	16395
Qy	855	GCAGGACGGGTCCAGGTGTGAGCTGACATCCGTGGCTGGCTGTGGTGGATGCCGGGA	914
Db	16396	AAAGSATGACATTAAGCATTTCAATGGTGATCAAGTCAGCTGCTTTTGAAGATGAAGCAAA	16455
Qy	915	GTACTCGTGTGTGCGGCGAGGAGAGACCTTCAGCCACACTCTACTGTCAAGGGCCCTGCC	974
Db	16456	ATACATGTTTGAAGCTGAAGATAAGCACAAAGTGGCAAACTGATCATTTGAAGGATCCG	16515
Qy	975	TGCCAGATTCAAGAGATGTGAAAAA	1001
Db	16516	GCTCAAATTCCTCACCCCTCTCAAAGA	16542

RESULT 12

```

RES001 12
US-09-759-508B-1
; Sequence 1, Application US/09759508B
; Publication No. US20020182599A1
; GENERAL INFORMATION:
; APPLICANT: Fishman, Mark C.
; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
; FILE REFERENCE: 00786/381002
; CURRENT APPLICATION NUMBER: US/09/759,508B
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/175,787
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 81940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (133)..(80910)
; OTHER INFORMATION:
; US-09-759-508B-1

```

Query Match

5.7%; Score 57.4; DB 9; Length 81940;

Best Local Similarity 43.3%; Pred. No. 6.3e-06;
Matches 375; Conservative 0; Mismatches 486; Indels 6; Gaps 2;
QY 141 GCAGGAGGAGCCTCAGGCACGCTCACCATCAGGGCTCGCTGCCAGGTTTCATAGAAGA 200
Db 15585 GGAGCTGAGCTCACAGCAGCTGAGTCTTAGAGGCCGATCCCTACTCTCAGTGAA 15644
QY 201 TGTGAAAAACCCAGGAGGCCAGAGAAGGGCCACCGCTGTGCTGCAGTGTGAGCTGAACAG 260
Db 15645 ATTACATGACAAACTGCAGTGGAGAGGATGAGATTACTTTGAAAGTGTGAAGTGACAA 15704
QY 261 TGCAGCCCTCTGGAGTGAGAAAGGGTCTGAGACCTCAGAGATGGGACAGATACAG 320
Db 15705 AGATGTACCAGTGAATGTTCAAGATGGTGAAGAGATTGTCCCTTCACCCAAATATTC 15764
QY 321 CCTGAGGAGGAGCGGACTAAATGTGAGCTGCAGATTCGTGGCCCTGGCCATGGCAGACAC 380
Db 15765 TATCAGGAGATGGCTCGCGCGCATCTTAAANAATCAAAAAGCGGACCTTAAAGATAA 15824
QY 381 TGGGAGTACTCGTGTGCGGCGAGGAGGACCTCGGCTATGCTCACCCTCAGGGC 440
Db 15825 AGCGGAATATGTGTGAGCTGTGGCACAGACAAGACCAAGGCAATGTTACTGTGAGGC 15884
QY 441 TCTACCCATCAATTCACAGAGGCTGTGAGAACGAAGGCCACAGAGGGGCCAACAGC 500
Db 15885 TCGACTAATAGAAAGTGGAAAGCCCTCTGTACGAGTAGAGGTGTTTGTGTGTAACAGC 15944
QY 501 CGTGTGCGGTGTGAGCTGAGCAAGATGGCCCGCTGGAGTGGTGGAGGGCGCATG---A 557
Db 15945 CCACCTTGAAATGAACTTCTGAACCTGTGTAACCTGTATGTTACGCCCATGGAAAGTGA 16004
QY 558 GACCCTCAGAGATGAGACAGACAGCCTGAGGAGGACCGGGCCAGGCTGTGAGCTGCA 617
Db 16005 GCCTTTGACAGCTTCCCTGACTGTGAAATCAATCAGGATGGAAGAGCATATTTCTGAT 16064
QY 618 GATCCGGCCCTCGTGGCAGAGGACGCTGGGAGTACCTGTGCATGTGCGGGAAGGAG 677
Db 16065 CCTTCATAACTGTCTGAGTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGTCTAATGCCAA 16124
QY 678 GACCTCAGCCATGCTCAGCGTCAGGCGCATCGCTTCCAAGTTTCATAGAGGCTCTGAGGAA 737
Db 16125 ATCTCAGCCCAATCTGAAAGTGAAGAAATGGCTTCTATCTCATCACCTCTCAGTGA 16184
QY 738 TGAAGAGCCACAGAAGGGGACAGCGCCACGCTGTGAGTGTGAGTGAAGGAGGCGC--- 794
Db 16185 TGTAAAGTCTTCGAGAAAGATGAGGCTAAGTTTGAGTGTGAAGTATCCAGGAGGCCAA 16244
QY 795 ACCGGTGGAGTGGAGAAAGGSCATGAGACCTCAGAGATGGGACAGACACAGCTGAG 854
Db 16245 AACATTCGTTGGCTAAAGGAACCCAGGAATCAGAGTATGACAGATTTGAGCTTAT 16304
QY 855 GCAGACGGGTCCAGGTGTGAGCTGCAGATCCGTGGCTGTGTTGGTGGATGCGGGGA 914
Db 16305 AAAGATGGCACTAAGCAATCAATGGTATCAAGTCAGCTGCTTTTGAAGATGAAGCAA 16364
QY 915 GTACTGTGTGTGCGGGAGGAGAGCACTCAGCCACATCTACTGTGAGGCGCCCTGCC 974
Db 16365 ATACATGTTTGAAGCTGAAGATAAGCACACAAGTGGCAAACTGATCATTTGAAGGAATCCG 16424
QY 975 TGCCAGATTTCATAGAAGATGTGAAAA 1001
Db 16425 GCTCAATTCCTCACCCCTCTCAAGA 16451

RESULT 13
US-09-960-706-1092
; Sequence 1092, Application US/09560706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; TITLE OF INVENTION: Gene Expression Profiles
; FILE REFERENCE: 44921-5029-01US

; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1092
; LENGTH: 81940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X90568
US-09-960-706-1092

Query Match 5.7%; Score 57.4; DB 10; Length 81940;
Best Local Similarity 43.3%; Pred. No. 6.3e-06;
Matches 375; Conservative 0; Mismatches 486; Indels 6; Gaps 2;
QY 141 GCAGGAGGAGCCTCAGGCACGCTCACCATCAGGGCTCGCTGCCAGGTTTCATAGAAGA 200
Db 15585 GGAGCTGAGCTCACAGCAGCTGAAAGTCTTAGAGGCCGATCCCTACTCTCAGTGAA 15644
QY 201 TGTGAAAAACCCAGGAGGCCAGAGAAGGGGCCACCGCTGTGCTGCAGTGTGAGCTGAACAG 260
Db 15645 ATTACATGACAAACTGCAGTGGAGAGGATGAGATTACTTTGAAAGTGTGAAGTGACAA 15704
QY 261 TGCAGCCCTCTGGAGTGGAGAAAGGGTCTGAGACCTCAGAGATGGGACAGATACAG 320
Db 15705 AGATGTACCAGTGAATGTTCAAGATGGTGAAGAGATTGTCCCTTCACCCAAATATTC 15764
QY 321 CCTGAGGAGGAGCGGACTAAATGTGAGCTGCAGATTCGTGGCCCTGGCCATGGCAGACAC 380
Db 15765 TATCAGGAGATGGCTCGCGCGCATCTTAAANAATCAAAAAGGCGACCTTAAAGATAA 15824
QY 381 TGGGAGTACTCGTGTGCGGCGAGGAGGACCTCGGCTATGCTCACCCTCAGGGC 440
Db 15825 AGCGGAATATGTGTGACTGTGGCACAGACAAGCAAGGCAATGTTACTGTGAGGC 15884
QY 441 TCTACCCATCAAGTTCACAGAGGCTGTGAGAACGAAGGCCACAGAGGGGCCAACAGC 500
Db 15885 TCGACTAATAGAAAGTGGAAAGCCCTCTGTACGAGTAGAGGTGTTTGTGTGTAACAGC 15944
QY 501 CGTGTGCGGTGTGAGCTGAGCAAGATGGCCCGCTGGAGTGGTGGAGGGCGCATG---A 557
Db 15945 CCACCTTGAAATGAACTTCTGAACTGTGATGTTTCAGGCCATGGAAAGTGAAGGACA 16004
QY 558 GACCCTCAGAGATGAGACAGACAGCCTGAGGAGGACCGGGCCAGGCTGTGAGCTGCA 617
Db 16005 GCCTTTGACAGCTTCCCTGACTGTGAAATCAATGAGGATGGAAGAGCATATTTCTGAT 16064
QY 618 GATCCGGCCCTCGTGGCAGAGACCTCGGAGTACCTGTGCATGTGCGGGAAGGAGAG 677
Db 16065 CCTTCATAACTGTCTGAGTGGGTATGACAGGAGAGGTTTCTTCCAGGCTGTCTAATGCCAA 16124
QY 678 GACCTCAGCCATGCTCAGCGTCAGGCGCATCGCTTCCAAGTTTCATAGAGGCTCTGAGGAA 737
Db 16125 ATCTCAGCCCAATCTGAAAGTGAAGAAATGGCTTCTATCTCATCACCTCTCAGTGA 16184
QY 738 TGAAGAGCCACAGAAGGGGACAGCGCCACGCTGTGAGTGTGAGTGAAGGAGGCGC--- 794
Db 16185 TGTAAAGTCTTCGAGAAAGATGAGGCTAAGTTTGAGTGTGAAGTATCCAGGAGGCCAA 16244
QY 795 ACCGGTGGAGTGGAGAAAGGSCATGAGACCTCAGAGATGGGACAGACACAGCTGAG 854
Db 16245 AACATTCGTTGGCTAAAGGAACCCAGGAATCAGAGTATGACAGATTTGAGCTTAT 16304
QY 855 GCAGACGGGTCCAGGTGTGAGCTGCAGATCCGTGGCTGTGTTGGTGGATGCGGGGA 914
Db 16305 AAAGATGGCACTAAGCAATCAATGGTATCAAGTCAGCTGCTTTTGAAGATGAAGCAA 16364
QY 915 GTACTGTGTGTGCGGGAGGAGAGCACTCAGCCACATCTACTGTGAGGCGCCCTGCC 974
Db 16365 ATACATGTTTGAAGCTGAAGATAAGCACACAAGTGGCAAACTGATCATTTGAAGGAATCCG 16424
QY 975 TGCCAGATTTCATAGAAGATGTGAAAA 1001
Db 16425 GCTCAATTCCTCACCCCTCTCAAGA 16451


```

Qy 441 TCTACCCATCAAGTTCACAGAGGCTCTGAGGAACGAGAGGCCACAGAGGGGCAACAGC 500
Db 15885 TCGACTAATAGAGGTGAAAGCCCTCTGTACGGAGTAGAGGTGTTTGTGTGAAACAGC 15944
Qy 501 CGTGTGGGTGTGAGCTGAGCAGAGATGGCCCCGTGGAGTGTGTGGAAGGGGCATG---A 557
Db 15945 CCACCTTGAATTTGAACTTCTTGAACTGATGTTTACGGCCAGTGGGAAGCTGGAAGGACA 16004
Qy 558 GACCTCAGAGATGGAGACAGACACAGCCTGAGGACAGACGGGGCCAGGTGTGAGCTGCA 617
Db 16005 GCCTTTGACAGCTTCCCTGCTGCTGAAATCATTTGAGGATGGAAGAACATATTTCTGAT 16064
Qy 618 GATCCGGGGCTCTGSCAGAGGACGCTGGGGAGTACCTGTGCATGTGCGGAGGAGAG 677
Db 16065 CCTTCATAACTGTGAGCTGGGTATGACAGGAGAGGTTTCCCTCCAGGCTGCTAATGCCAA 16124
Qy 678 GACCTCAGCCATGCTCACCGTCAAGGCGCATGCGCTTCCAAGTTCCATAGAGGTTCTGAGAA 737
Db 16125 ATCTGCAGCCAACTGAAAGTGAAGAATTGCTCTTTATCTTCATCACACCTCTCAGTGA 16184
Qy 738 TGAAGAGCCACAGAGGGGACACGGCCACGCTGTGTGTGAGCTGAGCAAGGGCGC--- 794
Db 16185 TGTAAAGTCTTCGAGAAAGATGAGGCTTAAGTTTGAAGTGAAGTATCCAGGGAGCCCAA 16244
Qy 795 ACCGTGTGAGTGGAGGAAGGGGCATGAGACCTCAGAGATGGGACAGACACAGCCTGAG 854
Db 16245 AACATTCGGTTGGCTAAAGGAACCCAGGAATCACAGGTGATGACAGATTTGAGCTTAT 16304
Qy 855 GCAGAGCGGTCCAGGTGTGAGTGCAGATCCGTGGCTGGCTGTGTGGATGCCGGGA 914
Db 16305 AAAGGATGGCACTAAGCATTCAATGGTGATCAAGTCAGCTGCTTTTGAAGATGAAGCAA 16364
Qy 915 GTACTCGTGTGTGCGGGCAGGAGGACCTCAGCCACACTCACTGTGAGGGCCCTGCC 974
Db 16365 ATACATGTTTGAAGCTGAAGATAGACACAAAGTGGCAAACTGATCATTTGAAGGAATCCG 16424
Qy 975 TGCCAGATTCATAGAAGATGTGA AAAA 1001
Db 16425 GCTCAATTCCTACCCCTCTCAAGA 16451

```

Search completed: March 21, 2005, 16:07:39
 Job time : 633.052 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2005, 17:29:58 ; Search time 496 Seconds
(without alignments)
5309.984 Million cell updates/sec

Title: US-10-077-130-5

Perfect score: 41273

Sequence: 1 MDQPFSGARFLTRPKAFV.....NRREKRRALLYKRNLAQVR 7968

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1401741 seqs, 330541175 residues

Total number of hits satisfying chosen parameters: 1401741

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41273	100.0	7968	13	US-10-077-130-5
2	32134	77.9	6620	15	US-10-080-334-290
3	32134	77.9	6620	16	US-10-408-765A-2291
4	13710	33.2	2630	13	US-10-077-130-2
5	13528	32.8	2596	14	US-10-307-019-6
6	10519.5	25.5	4691	15	US-10-093-463-72
7	10506	25.5	4675	15	US-10-093-463-74
8	9834.5	23.8	2328	16	US-10-476-397-4
9	8423	20.4	1665	9	US-09-858-664A-2
10	8423	20.4	1665	14	US-10-274-978-2
11	8423	20.4	1665	15	US-10-697-263-2
12	8423	20.4	1665	15	US-10-921-168-2
13	8410	20.4	1665	15	US-10-415-011-22

14	8407	20.4	1618	15	US-10-182-243-46	Sequence 46, Appli
15	8403	20.4	1610	14	US-10-307-019-4	Sequence 4, Appli
16	7073	17.1	1351	14	US-10-307-019-1	Sequence 1, Appli
17	6619.5	16.0	1596	16	US-10-408-765A-992	Sequence 992, App
18	4533	11.0	871	14	US-10-307-019-7	Sequence 7, Appli
19	4529	11.0	871	15	US-10-311-034-20	Sequence 20, Appli
20	3580.5	8.7	26926	9	US-09-759-508B-2	Sequence 2, Appli
21	2629	6.4	4162	15	US-10-023-834-92	Sequence 92, Appli
22	2489	6.0	6642	15	US-10-369-493-5013	Sequence 5013, Ap
23	2447.5	5.9	548	14	US-10-307-019-8	Sequence 8, Appli
24	2415.5	5.9	548	14	US-10-307-019-9	Sequence 9, Appli
25	2324	5.6	3208	15	US-10-210-130-38	Sequence 38, Appli
26	2319	5.6	3268	15	US-10-379-381-2	Sequence 2, Appli
27	2270	5.5	3252	15	US-10-210-130-36	Sequence 36, Appli
28	2203	5.3	3262	15	US-10-379-381-4	Sequence 4, Appli
29	2135	5.2	2380	15	US-10-333-314-18	Sequence 18, Appli
30	2102	5.1	5635	15	US-10-451-168-78	Sequence 78, Appli
31	2085.5	5.1	3186	15	US-10-210-130-34	Sequence 34, Appli
32	2085	5.1	416	15	US-10-108-260A-4389	Sequence 4389, Ap
33	2072.5	5.0	5636	14	US-10-032-189-128	Sequence 128, App
34	2072.5	5.0	5636	15	US-10-120-801-72	Sequence 72, Appli
35	2072.5	5.0	5636	15	US-10-023-634-93	Sequence 93, Appli
36	2072.5	5.0	5636	16	US-10-408-765A-1895	Sequence 1895, Ap
37	2067	5.0	2231	15	US-10-379-381-5	Sequence 5, Appli
38	2067	5.0	2242	16	US-10-408-765A-793	Sequence 793, App
39	2048	5.0	5198	15	US-10-120-801-75	Sequence 75, Appli
40	2048	5.0	5198	15	US-10-369-493-6858	Sequence 6858, Ap
41	2048	5.0	5198	15	US-10-369-493-6860	Sequence 6860, Ap
42	2048	5.0	5198	15	US-10-346-863-34	Sequence 34, Appli
43	2045	5.0	390	15	US-10-425-114-37530	Sequence 37530, A
44	2043.5	5.0	5175	15	US-10-120-801-74	Sequence 74, Appli
45	2043.5	5.0	5175	15	US-10-369-493-6859	Sequence 6859, Ap

ALIGNMENTS

RESULT 1

US-10-077-130-5
; Sequence 5, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Thereof
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 7968
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-5

Query Match 100.0%; Score 41273; DB 13; Length 7968;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDQPFSGARFLTRPKAFVSVGKDATLSQIVGNPTPQVSWEKDQPVTAGARFLAQ 60
Db 1 MDQPFSGARFLTRPKAFVSVGKDATLSQIVGNPTPQVSWEKDQPVTAGARFLAQ 60
QY 61 DGDYRLTILDLALGDSQYVCRNNAIGEAFAAVGLQVDAEAAACAEQAPHPLLRPTSR 120
Db 61 DGDYRLTILDLALGDSQYVCRNNAIGEAFAAVGLQVDAEAAACAEQAPHPLLRPTSR 120
QY 121 VREGSEATFCRCVGSPPAVSWKDGRIEGEPDGPVRVRELGEASALRIEAAAPRGG 180

Db 121 VREGSEATFRRCRVGSPRAVWSKQGRRLGCEPDGPRVVRVEELGEASALRIRARPRDGG 180
Qy 181 TYEVRANPLGAASAAAALVDSADATASRPGCTSTAALLAHILORREANRAGAPASPP 240
Db 181 TYEVRANPLGAASAAAALVDSADATASRPGCTSTAALLAHILORREANRAGAPASPP 240
Qy 241 STGTRTCTVTGEGKHARLSCVVTGEPKPTVMKDGQLVTEGRRHVVEDAQENFVLKILF 300
Db 241 STGTRTCTVTGEGKHARLSCVVTGEPKPTVMKDGQLVTEGRRHVVEDAQENFVLKILF 300
Qy 301 CKOSDRGLYCTASNLVQGYSSVLVVRREPAPVFKRLDLEVRKESATFLCEVPQPS 360
Db 301 CKOSDRGLYCTASNLVQGYSSVLVVRREPAPVFKRLDLEVRKESATFLCEVPQPS 360
Qy 361 TEAAWFKEETRLWASAKYGIEEGCTERRLTVRNVSADDDAVYICETPEGSRTVAELAVOG 420
Db 361 TEAAWFKEETRLWASAKYGIEEGCTERRLTVRNVSADDDAVYICETPEGSRTVAELAVOG 420
Qy 421 NLLRKLPRKTAVRGDTAMFCVELAVPVGPHWLRNQEVEVAGGRVAISAEGRTRHTLTIS 480
Db 421 NLLRKLPRKTAVRGDTAMFCVELAVPVGPHWLRNQEVEVAGGRVAISAEGRTRHTLTIS 480
Qy 481 QCCELDVGQVAFWAGDCOTSTRECVSAPRKPPLQPPVDPVVKARMESSVILSWSPPHGE 540
Db 481 QCCELDVGQVAFWAGDCOTSTRECVSAPRKPPLQPPVDPVVKARMESSVILSWSPPHGE 540
Qy 541 RPVTIDGYLVEKKKLGTYTWRICHEAEWATBELTVADVAEENGFQFRVSALNSFGQSPY 600
Db 541 RPVTIDGYLVEKKKLGTYTWRICHEAEWATBELTVADVAEENGFQFRVSALNSFGQSPY 600
Qy 601 LEPPGTVHLAPKLVATPLKAVQAVEGGEVTFSDLTVASAGWFIDGQALKASSVYEIH 660
Db 601 LEPPGTVHLAPKLVATPLKAVQAVEGGEVTFSDLTVASAGWFIDGQALKASSVYEIH 660
Qy 661 CDRTRHTLTIREVPASLHGAQLKFVANGIESSIRMEVRAAPGLTANKPAAAAREVLARL 720
Db 661 CDRTRHTLTIREVPASLHGAQLKFVANGIESSIRMEVRAAPGLTANKPAAAAREVLARL 720
Qy 721 HEEAQLLAELSQAATAVTLKDGRTLSPGPKYEVQASAGRRVLLVRDVARDDAGLYECVS 780
Db 721 HEEAQLLAELSQAATAVTLKDGRTLSPGPKYEVQASAGRRVLLVRDVARDDAGLYECVS 780
Qy 781 RGRITAYQLSVQGLARFLHOKWAGSCVDVAGGPAQFCECTSEAHVHVHWYKDGMELGHS 840
Db 781 RGRITAYQLSVQGLARFLHOKWAGSCVDVAGGPAQFCECTSEAHVHVHWYKDGMELGHS 840
Qy 841 GERFLOEDVGTTRHLVAATVTTRQDEGTYSYCRVGEDSDVDFRLRVSEPKVVFPAKEQLARRKL 900
Db 841 GERFLOEDVGTTRHLVAATVTTRQDEGTYSYCRVGEDSDVDFRLRVSEPKVVFPAKEQLARRKL 900
Qy 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMBATGCTTRRLVVQAGQADAGEY 960
Db 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMBATGCTTRRLVVQAGQADAGEY 960
Qy 961 SCEAGGQRLSFHLVDYKBPVFAKDQVAHSEVQAEAGANATLSCEVAQAQAEVVMWYKDGK 1020
Db 961 SCEAGGQRLSFHLVDYKBPVFAKDQVAHSEVQAEAGANATLSCEVAQAQAEVVMWYKDGK 1020
Qy 1021 KLSSSLKVHVEAKGRRRLVWQAGKTDAGDYSCEARGQVRSFRLHITTEPQWFAKEQSV 1080
Db 1021 KLSSSLKVHVEAKGRRRLVWQAGKTDAGDYSCEARGQVRSFRLHITTEPQWFAKEQSV 1080
Qy 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVCMBATGCTTRRLVLVQAGKAD 1140
Db 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVCMBATGCTTRRLVLVQAGKAD 1140
Qy 1141 AGEYSCEAGGQVRSFHLHITTEPQWFAKEQSVHNEVQAEAGTAMLSCEVAQAQTEVTWY 1200
Db 1141 AGEYSCEAGGQVRSFHLHITTEPQWFAKEQSVHNEVQAEAGTAMLSCEVAQAQTEVTWY 1200
Qy 1201 KDGKLSSSSKVMEVKGCTTRRLVVQGVKADAGEYCEAGGQVRSFQHLHITTEPKAVFAK 1260
Db 1201 KDGKLSSSSKVMEVKGCTTRRLVVQGVKADAGEYCEAGGQVRSFQHLHITTEPKAVFAK 1260

Qy 1261 EQLVNEVRTEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEAGCMRQLVVQQA 1320
Db 1261 EQLVNEVRTEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEAGCMRQLVVQQA 1320
Qy 1321 GQADAGEYTCAGGQRLSFHLVDSEPKAVFAKEQLAHRKVOAEAGAIATLSCEVAQAQTE 1380
Db 1321 GQADAGEYTCAGGQRLSFHLVDSEPKAVFAKEQLAHRKVOAEAGAIATLSCEVAQAQTE 1380
Qy 1381 VTWYKDGKLSSSSKVMEAVGCTTRRLVVQACQADTGEYCEAGGQRLSFSLDVABPKV 1440
Db 1381 VTWYKDGKLSSSSKVMEAVGCTTRRLVVQACQADTGEYCEAGGQRLSFSLDVABPKV 1440
Qy 1441 VPAKEQPVHREVQAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVMEAVGCTTRRLV 1500
Db 1441 VPAKEQPVHREVQAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVMEAVGCTTRRLV 1500
Qy 1501 VQAGQADAGEYSCEAGSQRLSFHLHVAEPKAVFAKEQPASREVQAEAGTATLSCEVAQ 1560
Db 1501 VQAGQADAGEYSCEAGSQRLSFHLHVAEPKAVFAKEQPASREVQAEAGTATLSCEVAQ 1560
Qy 1561 AQTEVTWYKDGKLSSSSKVMEAVGCTTRRLVVQAEQADAGEYSCKAGDQRLSFHLHVA 1620
Db 1561 AQTEVTWYKDGKLSSSSKVMEAVGCTTRRLVVQAEQADAGEYSCKAGDQRLSFHLHVA 1620
Qy 1621 EPKVVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRVEAVGCT 1680
Db 1621 EPKVVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRVEAVGCT 1680
Qy 1681 RRLVVQAGQADAGEYSCEAGGQRLSFRLHVAELPEQISERPCCRREPLVVKHEHEDIILTA 1740
Db 1681 RRLVVQAGQADAGEYSCEAGGQRLSFRLHVAELPEQISERPCCRREPLVVKHEHEDIILTA 1740
Qy 1741 TLATPSAATVTWLKDGVEIIRRSKRHETASQGDTHTLTVHGAQVLDLSAIYSCRVAEGQDF 1800
Db 1741 TLATPSAATVTWLKDGVEIIRRSKRHETASQGDTHTLTVHGAQVLDLSAIYSCRVAEGQDF 1800
Qy 1801 PVQVEEVAAKFCRLLEPVCGLGTTVLACELSPACAEVVMRCGNTQPRVKGFPQWVAEG 1860
Db 1801 PVQVEEVAAKFCRLLEPVCGLGTTVLACELSPACAEVVMRCGNTQPRVKGFPQWVAEG 1860
Qy 1861 PVRSRLTVGLRABDAGEYVCESRDDHTSAQLTVSVPRVVKFMSGLSTVVAEAGEATFQC 1920
Db 1861 PVRSRLTVGLRABDAGEYVCESRDDHTSAQLTVSVPRVVKFMSGLSTVVAEAGEATFQC 1920
Qy 1921 VVSPSDVAVVWFRDGLALLOPSEKFAISQSGASHSLTISDLVLEDAQOITVEAEGASSAA 1980
Db 1921 VVSPSDVAVVWFRDGLALLOPSEKFAISQSGASHSLTISDLVLEDAQOITVEAEGASSAA 1980
Qy 1981 LRVREAPVLFKKKLEPQTVTEERSVTVLEVLTRPWPPELRWTRNATALAPGKNVEIHAEGA 2040
Db 1981 LRVREAPVLFKKKLEPQTVTEERSVTVLEVLTRPWPPELRWTRNATALAPGKNVEIHAEGA 2040
Qy 2041 RHRLVLHNVGFADRGFGCETPDDKTQAKLTVEMROVRLVIRGLOAVEAREQGTATMEVOL 2100
Db 2041 RHRLVLHNVGFADRGFGCETPDDKTQAKLTVEMROVRLVIRGLOAVEAREQGTATMEVOL 2100
Qy 2101 SHADVDSWTRDGLRFQOQPTCHILAVRGPWHTLTLSGLRPEDSGLMVFKAEGVHTSARLV 2160
Db 2101 SHADVDSWTRDGLRFQOQPTCHILAVRGPWHTLTLSGLRPEDSGLMVFKAEGVHTSARLV 2160
Qy 2161 VTLPVPSFRPLQDVVVTTEKEKVTLECELSRPNVDVRLKDGVELRAGKTMAIAAQGACR 2220
Db 2161 VTLPVPSFRPLQDVVVTTEKEKVTLECELSRPNVDVRLKDGVELRAGKTMAIAAQGACR 2220
Qy 2221 SLTIVRCEPADQGVVCDADHAQSSASVKVQCGRTTLYIRRVLAEDAGBIQFVAENAESR 2280
Db 2221 SLTIVRCEPADQGVVCDADHAQSSASVKVQCGRTTLYIRRVLAEDAGBIQFVAENAESR 2280
Qy 2281 AQLRVKELPVTILVRPLRDKIAMKEKIRGVLECCQVSRASQVRWFKGSQELQPGPKYELVSD 2340
Db 2281 AQLRVKELPVTILVRPLRDKIAMKEKIRGVLECCQVSRASQVRWFKGSQELQPGPKYELVSD 2340

QY	2341	GLYRKLIIISDVHAEDEDTYTCAGDVKTSAQFVBEQSIITIVRGLQDVTVMPEPAPWREC	2400
Db	2341	GLYRKLIIISDVHAEDEDTYTCAGDVKTSAQFVBEQSIITIVRGLQDVTVMPEPAPWREC	2400
QY	2401	ETSPSVPPKWLKTKTQLQAGNVGLRQEGTVHRLMLRRTCSWTGPHFTVGHKSRSSA	2460
Db	2401	ETSPSVPPKWLKTKTQLQAGNVGLRQEGTVHRLMLRRTCSWTGPHFTVGHKSRSSA	2460
QY	2461	RLVVSDDIPVLTPLPEKTBRELQSVLSCDPRPAPKAVQWYKDDTPLSPSKFKWMSLEG	2520
Db	2461	RLVVSDDIPVLTPLPEKTBRELQSVLSCDPRPAPKAVQWYKDDTPLSPSKFKWMSLEG	2520
QY	2521	QMAELRIIRLMPADAGVYRCQAGSAHSTEVTVAREVTVTCPLQDAEATEEGWASFSCE	2580
Db	2521	QMAELRIIRLMPADAGVYRCQAGSAHSTEVTVAREVTVTCPLQDAEATEEGWASFSCE	2580
QY	2581	LSHEDEEVEWSLNGMPLNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSILKVSTARS	2640
Db	2581	LSHEDEEVEWSLNGMPLNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSILKVSTARS	2640
QY	2641	EVVRKPVVFLKALDDLSAEBRGTLALQCEVSDPEAHVVRKDGVLGSPDKYDFLHTAGT	2700
Db	2641	EVVRKPVVFLKALDDLSAEBRGTLALQCEVSDPEAHVVRKDGVLGSPDKYDFLHTAGT	2700
QY	2701	RGLVVHVSYPDAGLYTCHVSEETRARVRVHDLHVGITKRLKTMVLEGECSPECVLS	2760
Db	2701	RGLVVHVSYPDAGLYTCHVSEETRARVRVHDLHVGITKRLKTMVLEGECSPECVLS	2760
QY	2761	HESASDPAMTVGGKTVGSSRFQATROGRKYILVVBREAPSDAGEVVSFVRLGTSKASL	2820
Db	2761	HESASDPAMTVGGKTVGSSRFQATROGRKYILVVBREAPSDAGEVVSFVRLGTSKASL	2820
QY	2821	IVRERPAAI IKPLEDQWVAPGEDVLRCELSSAGTPVHVLKDRKAI RKSQKYDVVCEGTM	2880
Db	2821	IVRERPAAI IKPLEDQWVAPGEDVLRCELSSAGTPVHVLKDRKAI RKSQKYDVVCEGTM	2880
QY	2881	AMLVIRGASLKDAGEYTCVEVASKTASHVVEEKANCFTTELNTLOVEKGTAVFTCKTE	2940
Db	2881	AMLVIRGASLKDAGEYTCVEVASKTASHVVEEKANCFTTELNTLOVEKGTAVFTCKTE	2940
QY	2941	HPAATVTVWRKGLLELRASGKQHPQSEGLTLRLTISALEKADSDTYTCIDIGQAQSLV	3000
Db	2941	HPAATVTVWRKGLLELRASGKQHPQSEGLTLRLTISALEKADSDTYTCIDIGQAQSLV	3000
QY	3001	QGRRHIIIEDLVQVQSSATFRCTISPANYEPVHMFPLDKTPHLANELNEIDAQPGY	3060
Db	3001	QGRRHIIIEDLVQVQSSATFRCTISPANYEPVHMFPLDKTPHLANELNEIDAQPGY	3060
QY	3061	HVLTLROLALKDSGTIYFAGDORASALRVTEKPSFSRELTDATITTEGEDTLVLCETS	3120
Db	3061	HVLTLROLALKDSGTIYFAGDORASALRVTEKPSFSRELTDATITTEGEDTLVLCETS	3120
QY	3121	TCDIPMCWTCKGKTLRGSAQCQSHQRAQLLITGATLQDSGRYKCEAGGACSSSIVRV	3180
Db	3121	TCDIPMCWTCKGKTLRGSAQCQSHQRAQLLITGATLQDSGRYKCEAGGACSSSIVRV	3180
QY	3181	HARVPRFOEALKOLEVLEGGAAATLRCVLSVAAPVKWCYGNVLRPDKYSLRQEGAMLE	3240
Db	3181	HARVPRFOEALKOLEVLEGGAAATLRCVLSVAAPVKWCYGNVLRPDKYSLRQEGAMLE	3240
QY	3241	LWVRLNRQDSGRYSCSGDQDTTSATLTATPAQFICKLRNKEATEGATATLRCELSKT	3300
Db	3241	LWVRLNRQDSGRYSCSGDQDTTSATLTATPAQFICKLRNKEATEGATATLRCELSKT	3300
QY	3301	APVEWRKGSSETLRDGDRLRQDGMCELOIRGLAMVDAAEYSCVCGEERTSASITIRPM	3360
Db	3301	APVEWRKGSSETLRDGDRLRQDGMCELOIRGLAMVDAAEYSCVCGEERTSASITIRPM	3360
QY	3361	PAHFIGRLRHQESIEGATATLRCELSKAAPVEWRKRESLRDGDRLRQDGMCELOIC	3420
Db	3361	PAHFIGRLRHQESIEGATATLRCELSKAAPVEWRKRESLRDGDRLRQDGMCELOIC	3420
QY	3421	GLAVADAGEYSCVCGEERTSATLTIVKALPAKTEGLRNEEAVEGATAMLMWCELSKVAPVE	3480

Db	3421	GLAVADAGEYSCVCGEERTSATLTIVKALPAKTEGLRNEEAVEGATAMLMWCELSKVAPVE	3480
QY	3481	WRKGPENLRDGDRLRQEGTRCELQICGLAMADAGEYLCVCGQERTSATLTIRALPARF	3540
Db	3481	WRKGPENLRDGDRLRQEGTRCELQICGLAMADAGEYLCVCGQERTSATLTIRALPARF	3540
QY	3541	IEDVKNQBARREGATAVLOCELNSAAPVEWRKGSSETLRDGDRLRQDGTCKELQIRGLAM	3600
Db	3541	IEDVKNQBARREGATAVLOCELNSAAPVEWRKGSSETLRDGDRLRQDGTCKELQIRGLAM	3600
QY	3601	ADTGEYSCVCGQERTSAMLTVRALPIKTEGLRNEEATEGATAVLRCELSKAAPVEWRKGS	3660
Db	3601	ADTGEYSCVCGQERTSAMLTVRALPIKTEGLRNEEATEGATAVLRCELSKAAPVEWRKGS	3660
QY	3661	HETLRDGDRLRQDGMCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSKPIEGL	3720
Db	3661	HETLRDGDRLRQDGMCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSKPIEGL	3720
QY	3721	RNEEATEGDTATLWCELSKAAPVEWRKGSSETLRDGDRLRQDGSRCCELQIRGLAVVDAG	3780
Db	3721	RNEEATEGDTATLWCELSKAAPVEWRKGSSETLRDGDRLRQDGSRCCELQIRGLAVVDAG	3780
QY	3781	EYSCVCGQERTSATLTIVRALPARFIEDVKNQBARREGATAVLQCELSKAAPVEWRKGSSETL	3840
Db	3781	EYSCVCGQERTSATLTIVRALPARFIEDVKNQBARREGATAVLQCELSKAAPVEWRKGSSETL	3840
QY	3841	RGGDRYSLRQDGTTRCELQIHGLSVADTGEYSCVCGQERTSATLTIVRAPQVPVPREPLOSILQ	3900
Db	3841	RGGDRYSLRQDGTTRCELQIHGLSVADTGEYSCVCGQERTSATLTIVRAPQVPVPREPLOSILQ	3900
QY	3901	ABEGSTATLQCELSBPTATVWVSKGLOLQANGREPRLQGTAEVLQDLOREDTGEYT	3960
Db	3901	ABEGSTATLQCELSBPTATVWVSKGLOLQANGREPRLQGTAEVLQDLOREDTGEYT	3960
QY	3961	CTCGSQATSATLTIVTAAPVRFLELQHOEVDEGGTAHLCCELSRAGASVENEKGSILQPLP	4020
Db	3961	CTCGSQATSATLTIVTAAPVRFLELQHOEVDEGGTAHLCCELSRAGASVENEKGSILQPLP	4020
QY	4021	CARYQMVODGAAAEILLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRLQSLQEG	4080
Db	4021	CARYQMVODGAAAEILLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRLQSLQEG	4080
QY	4081	TGDIARLCCQLSDABSGAVQWLKEGVELHAGPKYEMRSQGNATRELLIHOLEBAKTGEYA	4140
Db	4081	TGDIARLCCQLSDABSGAVQWLKEGVELHAGPKYEMRSQGNATRELLIHOLEBAKTGEYA	4140
QY	4141	CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADDEVEFSCEVSRAGATGVQWCLQGLPLQ	4200
Db	4141	CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADDEVEFSCEVSRAGATGVQWCLQGLPLQ	4200
QY	4201	SNEVTEVAVRDRGRIHTLRKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPLQDVQ	4260
Db	4201	SNEVTEVAVRDRGRIHTLRKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPLQDVQ	4260
QY	4261	LSSEGODASFCQRLSRASQBARWALGGVPLQANENMDITVBOGTLHLHLTHKVTLEDACT	4320
Db	4261	LSSEGODASFCQRLSRASQBARWALGGVPLQANENMDITVBOGTLHLHLTHKVTLEDACT	4320
QY	4321	VSFHVGTCSSEAQLKVTAKNTVVRGLENVEALEGGEALFECCOLSOPEVAHAHTWLLDDBV	4380
Db	4321	VSFHVGTCSSEAQLKVTAKNTVVRGLENVEALEGGEALFECCOLSOPEVAHAHTWLLDDBV	4380
QY	4381	RTSENAEYVFPENGLRHLLKLRNLRPQDSRCVTFLAGDMVTSAFITVRGWRLEILEPLKN	4440
Db	4381	RTSENAEYVFPENGLRHLLKLRNLRPQDSRCVTFLAGDMVTSAFITVRGWRLEILEPLKN	4440
QY	4441	AAVRAGAQRFTCTLSEAVPVGEASWYINGAAVQDDSDWTVDAGSQAALLLRSAQPHH	4500
Db	4441	AAVRAGAQRFTCTLSEAVPVGEASWYINGAAVQDDSDWTVDAGSQAALLLRSAQPHH	4500
QY	4501	AGEVTFACEDAVASARLTVLGLPDPPEAEVVAHSSHTVTLSSWAPMDSGGGLCGYRVE	4560

Db 4501 AGEVTFACRDAVASARLTVLGLPDPEDAEVVAHSGSHTVTL5WAAPMSDGGGLCGYRVE 4560
QY 4561 VKEGATQWRLCHELVPGPECVVDGLAPGETYFRVAAGVPVGAPEPVHLPQTVRLAEP 4620
Db 4561 VKEGATQWRLCHELVPGPECVVDGLAPGETYFRVAAGVPVGAPEPVHLPQTVRLAEP 4620
QY 4621 KPVPPOPSAPESRQVAAGSDVLSLELVVAEAGEVIVHKGWRIQPGRFVVSQGRQQML 4680
Db 4621 KPVPPOPSAPESRQVAAGSDVLSLELVVAEAGEVIVHKGWRIQPGRFVVSQGRQQML 4680
QY 4681 VIKGFTAEDQGYHCGLAGOSTCPAAATFOVALSPASVDEAPQPSLPPEAAQEGDLHLW 4740
Db 4681 VIKGFTAEDQGYHCGLAGOSTCPAAATFOVALSPASVDEAPQPSLPPEAAQEGDLHLW 4740
QY 4741 EALARKRMSRPTLDSISELPEEDGORSORLPOEAEEVAPDLSSEGYSTADELARTGDADL 4800
Db 4741 EALARKRMSRPTLDSISELPEEDGORSORLPOEAEEVAPDLSSEGYSTADELARTGDADL 4800
QY 4801 SHTSSDDESRACTPSLVTVLTKKAGRPGTSPLASKVGAPAAPSVKPOQOQEPPLAAVRPPLG 4860
Db 4801 SHTSSDDESRACTPSLVTVLTKKAGRPGTSPLASKVGAPAAPSVKPOQOQEPPLAAVRPPLG 4860
QY 4861 DLSTKDLGDPMSDKAAVKLOAAFKGYKVRKEMKQEGPMFSHTFGDTEAQVGDALRLBVCV 4920
Db 4861 DLSTKDLGDPMSDKAAVKLOAAFKGYKVRKEMKQEGPMFSHTFGDTEAQVGDALRLBVCV 4920
QY 4921 VASKADVRARWLKDGVELTDGRHHIDQLGDGTC5LLIAGLDRADAGCYTCOVSNKFGQV 4980
Db 4921 VASKADVRARWLKDGVELTDGRHHIDQLGDGTC5LLIAGLDRADAGCYTCOVSNKFGQV 4980
QY 4981 THSACVVGSGSSEAESGG5G5BDDAFRAARLHRLFRTKSPARV5DEELFLSADEGPA 5040
Db 4981 THSACVVGSGSSEAESGG5G5BDDAFRAARLHRLFRTKSPARV5DEELFLSADEGPA 5040
QY 5041 EPEEPADWQTYREDEHFCIRFEALTEARQAVTRQEMPATLIGIVEIKLVEQGRPRVEM 5100
Db 5041 EPEEPADWQTYREDEHFCIRFEALTEARQAVTRQEMPATLIGIVEIKLVEQGRPRVEM 5100
QY 5101 CISKETPAVPVPEPLPSLLTSDAAPVFLTELQNEVDGYPV5FDCVVVTGQPMPSVRWF 5160
Db 5101 CISKETPAVPVPEPLPSLLTSDAAPVFLTELQNEVDGYPV5FDCVVVTGQPMPSVRWF 5160
QY 5161 XDKLLEEDDHYMINEDQGGHQLIITAVVPADMGYRCLAENSMGV5STKAEFLRVDLTS 5220
Db 5161 XDKLLEEDDHYMINEDQGGHQLIITAVVPADMGYRCLAENSMGV5STKAEFLRVDLTS 5220
QY 5221 TDYDTAADAATES5SYFSAQGYLSSREOQTESTTDEGQLPVQVEELRDLQVAPGTRLAKF 5280
Db 5221 TDYDTAADAATES5SYFSAQGYLSSREOQTESTTDEGQLPVQVEELRDLQVAPGTRLAKF 5280
QY 5281 QLKVKGYAPRLYWFKDGQPLTASAHIRMTGKKILHTLEIISVTREDSGQVAAVYSNANG 5340
Db 5281 QLKVKGYAPRLYWFKDGQPLTASAHIRMTGKKILHTLEIISVTREDSGQVAAVYSNANG 5340
QY 5341 AAY5SARLLVRGDPDEPKPADVHEQLVPPRMLERFTPKVKGSSITFFSVKVEGRVPV 5400
Db 5341 AAY5SARLLVRGDPDEPKPADVHEQLVPPRMLERFTPKVKGSSITFFSVKVEGRVPV 5400
QY 5401 TVHMLREAEARGVLTIGDTPGYTVASSAQHSLVLLDVGRHQGTYYTCIASNAAGQALC 5460
Db 5401 TVHMLREAEARGVLTIGDTPGYTVASSAQHSLVLLDVGRHQGTYYTCIASNAAGQALC 5460
QY 5461 SASLHVSGLPKVBEQKVEKALISTFLOGTTQAI5AQGLETFASPADLGGQKEEPLAAKE 5520
Db 5461 SASLHVSGLPKVBEQKVEKALISTFLOGTTQAI5AQGLETFASPADLGGQKEEPLAAKE 5520
QY 5521 ALGHLSLAEVGTETFLQKLT5QITEMWSAKITQAKLQVPGGSD5D5SKTP5ASPRHGRSR 5580
Db 5521 ALGHLSLAEVGTETFLQKLT5QITEMWSAKITQAKLQVPGGSD5D5SKTP5ASPRHGRSR 5580
QY 5581 PSS5TOESS5SE5D5GD5ARGE5FDI5VVV7ADY5PLCAEODAIT5REGQVVEVLDA5HPLRW 5640
Db 5581 PSS5TOESS5SE5D5GD5ARGE5FDI5VVV7ADY5PLCAEODAIT5REGQVVEVLDA5HPLRW 5640

QY 5641 LVRTKPTKSSPSRQGW5PAYLDRRLKLS5PWGA5A5E5PE5F5CE5V5E5DE5Y5KARL5SV5IQE 5700
Db 5641 LVRTKPTKSSPSRQGW5PAYLDRRLKLS5PWGA5A5E5PE5F5CE5V5E5DE5Y5KARL5SV5IQE 5700
QY 5701 LLS5EQAFVE5ELO5FLQ5SHHLO5HL5RC5PHV5PIA5VAG5Q5KAVI5FRNVRDI5GRF5HSS5FLQ5ELQO 5760
Db 5701 LLS5EQAFVE5ELO5FLQ5SHHLO5HL5RC5PHV5PIA5VAG5Q5KAVI5FRNVRDI5GRF5HSS5FLQ5ELQO 5760
QY 5761 CDTDD5VAM5CFI5KQAA5FEQ5YLE5FLV5GRV5Q5ES5VV5V5TAI5O5EFY5KY5ABE5ALL5AG5D5P5OP 5820
Db 5761 CDTDD5VAM5CFI5KQAA5FEQ5YLE5FLV5GRV5Q5ES5VV5V5TAI5O5EFY5KY5ABE5ALL5AG5D5P5OP 5820
QY 5821 PPPPL5QHY5LEQ5P5ERV5Q5YQ5ALL5KEL5IRN5KARN5QNC5ALL5EQ5AV5V5SAL5PQ5AEN5KL5HV 5880
Db 5821 PPPPL5QHY5LEQ5P5ERV5Q5YQ5ALL5KEL5IRN5KARN5QNC5ALL5EQ5AV5V5SAL5PQ5AEN5KL5HV 5880
QY 5881 SLMENY5PGL5E5PIR5Q5GHFI5VW5GAP5ARM5PW5K5GHN5RH5VFI5FRN5HLVI5CK5PR5D5RT 5940
Db 5881 SLMENY5PGL5E5PIR5Q5GHFI5VW5GAP5ARM5PW5K5GHN5RH5VFI5FRN5HLVI5CK5PR5D5RT 5940
QY 5941 DTVS5VFRN5M5KL5S5IDL5NDQ5VEG5DD5RA5FV5W5Q5RED5SV5RY5LLQ5ART5AI5TK5SW5VK5EIC 6000
Db 5941 DTVS5VFRN5M5KL5S5IDL5NDQ5VEG5DD5RA5FV5W5Q5RED5SV5RY5LLQ5ART5AI5TK5SW5VK5EIC 6000
QY 6001 GIQORLAL5PV5WR5P5DF5E5EL5AD5CT5A5EL5GET5VK5L5AC5RV5TG5TP5K5P5VI5SW5Y5K5DG5K5AV5Q5VD5PHH 6060
Db 6001 GIQORLAL5PV5WR5P5DF5E5EL5AD5CT5A5EL5GET5VK5L5AC5RV5TG5TP5K5P5VI5SW5Y5K5DG5K5AV5Q5VD5PHH 6060
QY 6061 ILI5BD5PG5CAL5I5D5SL5TG5VD5SG5Q5Y5M5C5FA5A5AG5NC5ST5LG5K5IL5V5Q5VP5FR5F5V5N5K5VR5AS5PFV 6120
Db 6061 ILI5BD5PG5CAL5I5D5SL5TG5VD5SG5Q5Y5M5C5FA5A5AG5NC5ST5LG5K5IL5V5Q5VP5FR5F5V5N5K5VR5AS5PFV 6120
QY 6121 EGE5DAQ5F5CT5T5LEG5AP5Q5I5RW5Y5K5DG5ALL5T5CN5K5F5QT5SE5PS5GL5LV5IRA5SK5ED5GLY 6180
Db 6121 EGE5DAQ5F5CT5T5LEG5AP5Q5I5RW5Y5K5DG5ALL5T5CN5K5F5QT5SE5PS5GL5LV5IRA5SK5ED5GLY 6180
QY 6181 EC5EL5VN5RL5GS5ARA5AE5L5RI5Q5SP5ML5QA5O5EQ5CH5RE5Q5L5VA5VED5T5TL5ER5AD5Q5E5V5T5VL5K5RL5LG 6240
Db 6181 EC5EL5VN5RL5GS5ARA5AE5L5RI5Q5SP5ML5QA5O5EQ5CH5RE5Q5L5VA5VED5T5TL5ER5AD5Q5E5V5T5VL5K5RL5LG 6240
QY 6241 PKA5P5ST5G5DL5TG5P5C5PR5GAP5AL5Q5ET5QS5Q5PV5T5GT5SE5AP5V5PR5V5P5O5PL5L5H5EG5PE5Q5B5PE 6300
Db 6241 PKA5P5ST5G5DL5TG5P5C5PR5GAP5AL5Q5ET5QS5Q5PV5T5GT5SE5AP5V5PR5V5P5O5PL5L5H5EG5PE5Q5B5PE 6300
QY 6301 AIA5RA5Q5E5W5TV5PIR5ME5GA5WP5GAG5TG5ELL5W5DV5H5SV5RE5T5T5Q5T5Y5Q5A5ID5T5HT5AR5PS5MQ 6360
Db 6301 AIA5RA5Q5E5W5TV5PIR5ME5GA5WP5GAG5TG5ELL5W5DV5H5SV5RE5T5T5Q5T5Y5Q5A5ID5T5HT5AR5PS5MQ 6360
QY 6361 VTI5ED5VQA5QT5G5TA5Q5FAI5IEG5DP5Q5PS5VT5W5Y5K5D5SV5QL5VD5STR5LS5Q5Q5EG5T5Y5SL5VR5HA 6420
Db 6361 VTI5ED5VQA5QT5G5TA5Q5FAI5IEG5DP5Q5PS5VT5W5Y5K5D5SV5QL5VD5STR5LS5Q5Q5EG5T5Y5SL5VR5HA 6420
QY 6421 SKD5AGV5T5CLA5QNT5GG5VLC5KAE5LL5VLC5G5D5NE5P5SE5K5SH5RR5KL5HS5F5VE5V5KE5E5T5GR5GV5FG 6480
Db 6421 SKD5AGV5T5CLA5QNT5GG5VLC5KAE5LL5VLC5G5D5NE5P5SE5K5SH5RR5KL5HS5F5VE5V5KE5E5T5GR5GV5FG 6480
QY 6481 FV5KE5VQ5HK5N5K5IL5CA5AK5FI5PL5RS5TRA5Q5A5VER5DI5LA5L5SH5PL5VT5GL5LD5Q5P5ET5R5K5TL5LI 6540
Db 6481 FV5KE5VQ5HK5N5K5IL5CA5AK5FI5PL5RS5TRA5Q5A5VER5DI5LA5L5SH5PL5VT5GL5LD5Q5P5ET5R5K5TL5LI 6540
QY 6541 LEL5CS5B5EL5DL5R5Y5RG5V5T5E5AE5VK5VI5Q5QL5VE5GL5H5L5H5SG5VL5H5LDI5K5PS5NI5LM5V5H5PAR 6600
Db 6541 LEL5CS5B5EL5DL5R5Y5RG5V5T5E5AE5VK5VI5Q5QL5VE5GL5H5L5H5SG5VL5H5LDI5K5PS5NI5LM5V5H5PAR 6600
QY 6601 EDI5KI5D5FG5PA5QNT5P5A5EL5Q5F5S5Q5GS5PE5F5V5PEI5IQ5NP5V5E5AS5DI5W5M5G5V5IS5L5T5CS 6660
Db 6601 EDI5KI5D5FG5PA5QNT5P5A5EL5Q5F5S5Q5GS5PE5F5V5PEI5IQ5NP5V5E5AS5DI5W5M5G5V5IS5L5T5CS 6660
QY 6661 SP5PAGES5DR5AT5LL5N5VL5GR5V5SW5SP5MA5HL5SED5AK5DFI5K5AT5LQ5RAP5Q5AR5P5AA5O5CL5SH5PW 6720
Db 6661 SP5PAGES5DR5AT5LL5N5VL5GR5V5SW5SP5MA5HL5SED5AK5DFI5K5AT5LQ5RAP5Q5AR5P5AA5O5CL5SH5PW 6720


```
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 290
; LENGTH: 6620
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-334-290

Query Match      77.9%; Score 32134; DB 15; Length 6620;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 6250; Conservative 15; Mismatches 57; Indels 48; Gaps 9;

QY 1 MDQPFSGAPRLTRPKAFVSVGKDATLSQOIVGNPTPOVSWEXDQOPVTAGAFRLAQ 60
DB 1 MDQPFSGAPRLTRPKAFVSVGKDATLSQOIVGNPTPOVSWEXDQOPVTAGAFRLAQ 60
QY 61 DGDLYRLTILDALGDSGOYVCRARNAIGEAFAAVGLQVDAEAAAEQAPHLRPTISR 120
DB 61 DGDLYRLTILDALGDSGOYVCRARNAIGEAFAAVGLQVDAEAAAEQAPHLRPTISR 120
QY 121 VREGSEATFRCRVGSPRPVSWSDGRRLGEPDGRVVRVBEKGASALIRTRAAPRDGG 180
DB 121 VREGSEATFRCRVGSPRPVSWSDGRRLGEPDGRVVRVBEKGASALIRTRAAPRDGG 180
QY 181 TYEVRAENPLGAASAAALVDSDAADTASRPGTSTAALLAHLQRRREARAEAPASPP 240
DB 181 TYEVRAENPLGAASAAALVDSDAADTASRPGTSTAALLAHLQRRREARAEAPASPP 240
QY 241 STGTRTCTVTEGKHARLSCTVTGEPKPKETVMKKDQQLVTEGRRHVYVEDAQENFVLKILF 300
DB 241 STGTRTCTVTEGKHARLSCTVTGEPKPKETVMKKDQQLVTEGRRHVYVEDAQENFVLKILF 300
QY 301 CKQDRGLYCTASNLVGGTYSVLVVRREPAPVPPKKRLQDLVREKESATFLCEVPQPS 360
DB 301 CKQDRGLYCTASNLVGGTYSVLVVRREPAPVPPKKRLQDLVREKESATFLCEVPQPS 360
QY 361 TEAAWFKEETRLWASAKYIEEGTERRLTVRNVSDADDVYICETPEGSRVTAELAVQG 420
DB 361 TEAAWFKEETRLWASAKYIEEGTERRLTVRNVSDADDVYICETPEGSRVTAELAVQG 420
QY 421 NLLRKLPRKTAVRVGDTFAMFCVELAVPVGVVHWRNQEVVAGGRVAISASGTRHTLTIS 480
DB 421 NLLRKLPRKTAVRVGDTFAMFCVELAVPVGVVHWRNQEVVAGGRVAISASGTRHTLTIS 480
QY 481 QCCLEDVQVAFMAGDCQTSRFCVSAPRKPPLOPPVDVVKARMESSVILSWSPPHGE 540
DB 481 QCCLEDVQVAFMAGDCQTSRFCVSAPRKPPLOPPVDVVKARMESSVILSWSPPHGE 540
QY 541 RPVTIDGYLVEKKLGTWTWIRCHEAEVWVATPELTVADVAEENFQPRVSAINLFGOSPY 600
DB 541 RPVTIDGYLVEKKLGTWTWIRCHEAEVWVATPELTVADVAEENFQPRVSAINLFGOSPY 600
QY 601 LEFPGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDTLTVASAGWFLDQALKASSVYIEH 660
DB 601 LEFPGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDTLTVASAGWFLDQALKASSVYIEH 660
QY 661 CDRTRHTLTIREVPASLHGAQLKPVANGIESSIRMEVRAAAGLGTANKPAAAAAEVLARL 720
DB 661 CDRTRHTLTIREVPASLHGAQLKPVANGIESSIRMEVRAAAGLGTANKPAAAAAEVLARL 720
QY 721 HEEAQLLAELSDQAAAVTWLKDGRTLSPGPKYEVQASAGRVLLVRDVARDDAGLYECVS 780
DB 721 HEEAQLLAELSDQAAAVTWLKDGRTLSPGPKYEVQASAGRVLLVRDVARDDAGLYECVS 780
QY 781 RGGRIAYQLSVQGLARFLHKDMAGSCVDVAGGPAQPECETSEARVHVHVKDGMELGHS 840
DB 781 RGGRIAYQLSVQGLARFLHKDMAGSCVDVAGGPAQPECETSEARVHVHVKDGMELGHS 840
QY 841 GERFLOEDVGTTHRLVAANTVTRODEGTVSCRVGSDSVDFRLRVSEPKVVFVFAKEQLARRKL 900
DB 841 GERFLOEDVGTTHRLVAANTVTRODEGTVSCRVGSDSVDFRLRVSEPKVVFVFAKEQLARRKL 900
```

```
DB 841 GERFLOEDVGTTHRLVAANTVTRODEGTVSCRVGSDSVDFRLRVSEPKVVFVFAKEQLARRKL 900
QY 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVCMEATGCTTRRLVVQAGQADAGEY 960
DB 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVCMEATGCTTRRLVVQAGQADAGEY 960
QY 961 SCEAGGQRLSPHLDVKEPKVVFVFAKDQVAHSVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
DB 961 SCEAGGQRLSPHLDVKEPKVVFVFAKDQVAHSVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
QY 1021 KLSSSLKHVBEAKCRRRLVVQAGKTADAGYSCAERGQVSRFLHITTEPROMFAKEQSV 1080
DB 1021 KLSSSLKHVBEAKCRRRLVVQAGKTADAGYSCAERGQVSRFLHITTEPROMFAKEQSV 1080
QY 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSSKVCMEVKGCTTRRLVLVQAGKAD 1140
DB 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSSKVCMEVKGCTTRRLVLVQAGKAD 1140
QY 1141 AGEYSCAAGGQVSRFLHITTEPKGVFAKEQSVHNEVQAEAGTATMLSCVAQAQTEVTWY 1200
DB 1141 AGEYSCAAGGQVSRFLHITTEPKGVFAKEQSVHNEVQAEAGTATMLSCVAQAQTEVTWY 1200
QY 1201 KDGKLSKSSSKVRMEVKGCTTRRLVVQVQKADAGEYSCAAGGQVSRFLHITTEPKAVPAK 1260
DB 1201 KDGKLSKSSSKVRMEVKGCTTRRLVVQVQKADAGEYSCAAGGQVSRFLHITTEPKAVPAK 1260
QY 1261 EQLVHNEVTRTEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAAAGCMROLVVQQA 1320
DB 1261 EQLVHNEVTRTEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAAAGCMROLVVQQA 1320
QY 1321 GQADAGEYTCCEAGGQRLSPHLDVSEPKAVFAKEQLAHRKQAEAGATATLSCEVAQAQTE 1380
DB 1321 GQADAGEYTCCEAGGQRLSPHLDVSEPKAVFAKEQLAHRKQAEAGATATLSCEVAQAQTE 1380
QY 1381 VTYWYKDGKLSKSSSKVRMEVKGCTTRRLVVQQAQADTGEYSCAAGGQRLSPFLDVABPKV 1440
DB 1381 VTYWYKDGKLSKSSSKVRMEVKGCTTRRLVVQQAQADTGEYSCAAGGQRLSPFLDVABPKV 1440
QY 1441 VFAKEQPVHREVOAQAGASTTSLCEVAQAQTEVTWYKDGKLSKSSSKVRMEVKGCTTRRLV 1500
DB 1441 VFAKEQPVHREVOAQAGASTTSLCEVAQAQTEVTWYKDGKLSKSSSKVRMEVKGCTTRRLV 1500
QY 1501 VQQAQADAGEYSCAAGGQRLSPHLHVABPKAVFAKEQPAQASREVQAEAGTATLSCEVAQ 1560
DB 1501 VQQAQADAGEYSCAAGGQRLSPHLHVABPKAVFAKEQPAQASREVQAEAGTATLSCEVAQ 1560
QY 1561 AQTEVTWYKDGKLSKSSSKVRMEVKGCTTRRLVVQEAQADAGEYSCAAGGQRLSPHLHVA 1620
DB 1561 AQTEVTWYKDGKLSKSSSKVRMEVKGCTTRRLVVQEAQADAGEYSCAAGGQRLSPHLHVA 1620
QY 1621 EPKVVFVFAKEQPAHREVOAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRVEAVGCT 1680
DB 1621 EPKVVFVFAKEQPAHREVOAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRVEAVGCT 1680
QY 1681 RRLVVQQAQADAGEYSCAAGGQRLSPHLHVAELEPOISERPCCRREPLVVKHEHDIILTA 1740
DB 1681 RRLVVQQAQADAGEYSCAAGGQRLSPHLHVAELEPOISERPCCRREPLVVKHEHDIILTA 1740
QY 1741 TLATPSSAATVTLKDGVEIIRSRKHETASQGDTHTLTVHGAQVLDLSAIYSCRVAEQDGF 1800
DB 1741 TLATPSSAATVTLKDGVEIIRSRKHETASQGDTHTLTVHGAQVLDLSAIYSCRVAEQDGF 1800
QY 1801 PVQVEEVAAKFCRLLEPVCCELGGTTLTACELSPACAEVVMRCNTOPRVGKRQFQVABG 1860
DB 1801 PVQVEEVAAKFCRLLEPVCCELGGTTLTACELSPACAEVVMRCNTOPRVGKRQFQVABG 1860
QY 1861 PVRSLTVGLRAEDAGEYVCESRDDHTSAQLTVSVPRVVKFMSGSLSTVVAEBGGEATFQC 1920
DB 1861 PVRSLTVGLRAEDAGEYVCESRDDHTSAQLTVSVPRVVKFMSGSLSTVVAEBGGEATFQC 1920
QY 1921 VVSPSDVAVVMFRDGLLOPSEKFAISQSGASHSLTISDLVLEDAQOITVVEAGASSAA 1980
DB 1921 VVSPSDVAVVMFRDGLLOPSEKFAISQSGASHSLTISDLVLEDAQOITVVEAGASSAA 1980
```

QY 1981 LRVRAPVLFKKLBPQTVBERSSVTLEVELTRPMPPELRWTRNATAPAGKNVEIHABGA 2040
DB 1981 LRVRAPVLFKKLBPQTVBERSSVTLEVELTRPMPPELRWTRNATAPAGKNVEIHABGA 2040
QY 2041 RHRLVLHNVGADRGFFGCEPDDKTQAKLTVMEMQVRLVRGLQVAEAREQGTATMEVOL 2100
DB 2041 RHRLVLHNVGADRGFFGCEPDDKTQAKLTVMEMQVRLVRGLQVAEAREQGTATMEVOL 2100
QY 2101 SHADVDGSWTRDGLRFQOQPTCHLAVRGPMTHTLTLSGLRPEDSGMLWFKAEGVHTSARLV 2160
DB 2101 SHADVDGSWTRDGLRFQOQPTCHLAVRGPMTHTLTLSGLRPEDSGMLWFKAEGVHTSARLV 2160
QY 2161 VTLEPVSFSRPLQDVVTTEKEKVTLECELSRPNVDVRLKOGVELRAGKTWAIQAQACR 2220
DB 2161 VTLEPVSFSRPLQDVVTTEKEKVTLECELSRPNVDVRLKOGVELRAGKTWAIQAQACR 2220
QY 2221 SLTIYRCFADQGVVCDADHAQSSASVKVQGRVTTLTYRRVLADAGEIQFVAENAESR 2280
DB 2221 SLTIYRCFADQGVVCDADHAQSSASVKVQGRVTTLTYRRVLADAGEIQFVAENAESR 2280
QY 2281 AOLRVKELPVTLRPLDKIAMEXHRGVLEQVSRASQVRFKGSQELQPGPKYELVSD 2340
DB 2281 AOLRVKELPVTLRPLDKIAMEXHRGVLEQVSRASQVRFKGSQELQPGPKYELVSD 2340
QY 2341 GLYRKLIIISDVHAEDEDTYTCDAGDVKTSAQPFVBEQSIITVIRGLQDVTVMEPAPAWFEC 2400
DB 2341 GLYRKLIIISDVHAEDEDTYTCDAGDVKTSAQPFVBEQSIITVIRGLQDVTVMEPAPAWFEC 2400
QY 2401 ETSIPSPVPPKWLKGTVLQAGNVGLRQEGVHRLMLRRTCTMTGPHFTVVGKSRSSA 2460
DB 2401 ETSIPSPVPPKWLKGTVLQAGNVGLRQEGVHRLMLRRTCTMTGPHFTVVGKSRSSA 2460
QY 2461 RLIVSDIIPVLTPLPEKTRGRELQSVLSCDFRPAPKAVQVQKDTPLSPSEKFKWLSLEG 2520
DB 2461 RLIVSDIIPVLTPLPEKTRGRELQSVLSCDFRPAPKAVQVQKDTPLSPSEKFKWLSLEG 2520
QY 2521 QMAELRIILRLMPADAGVTRCQAGSAHSSTEVTVAREVTVTGLODAEATBEGWASFSCE 2580
DB 2521 QMAELRIILRLMPADAGVTRCQAGSAHSSTEVTVAREVTVTGLODAEATBEGWASFSCE 2580
QY 2581 LSHDEEVEWSLNGMPLVNDSPFHEISHKGRHRTLVLKSIQRADAGIVRASSIKVSTARSAL 2640
DB 2581 LSHDEEVEWSLNGMPLVNDSPFHEISHKGRHRTLVLKSIQRADAGIVRASSIKVSTARSAL 2640
QY 2641 EVRVKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRWKGQVQLGPFSDKYDFLHTAGT 2700
DB 2641 EVRVKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRWKGQVQLGPFSDKYDFLHTAGT 2700
QY 2701 RGLVVHDSVPEDAGLYTCHVGSSETRARVRVHDHLHVGITKRLKTMVLEGSCESPCEVLS 2760
DB 2701 RGLVVHDSVPEDAGLYTCHVGSSETRARVRVHDHLHVGITKRLKTMVLEGSCESPCEVLS 2760
QY 2761 HESASDPAMWTVGGKTVGSSRFQATROGRKYILVVRQAAPSDAGEVVFVSRGLTSKASL 2820
DB 2761 HESASDPAMWTVGGKTVGSSRFQATROGRKYILVVRQAAPSDAGEVVFVSRGLTSKASL 2820
QY 2821 IVRERPAAIIPLEQDQWAPGEDVELRCELRAAGTPVHMLKDKRAIKRSQKYDVVCEGTM 2880
DB 2821 IVRERPAAIIPLEQDQWAPGEDVELRCELRAAGTPVHMLKDKRAIKRSQKYDVVCEGTM 2880
QY 2881 AMLVIRGASLKDAGEYTCVEBASKTASLHVEBKANCFTBELTNLQVBEKGTAVFTCKTE 2940
DB 2881 AMLVIRGASLKDAGEYTCVEBASKTASLHVEBKANCFTBELTNLQVBEKGTAVFTCKTE 2940
QY 2941 HPAATVTRKGLLELRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSAQLLV 3000
DB 2941 HPAATVTRKGLLELRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSAQLLV 3000
QY 3001 QGRRVHIIEDLVQDVGSSATFCRISPNAYEPVHFLDKTPLHANELNEIDAQPGGY 3060
DB 3001 QGRRVHIIEDLVQDVGSSATFCRISPNAYEPVHFLDKTPLHANELNEIDAQPGGY 3060

QY 3061 HVLTLRQALXKDSGTIYFEAGDQORASAAALRVTEKPSVPSRELTDTATITEGEDTLVLCBTS 3120
DB 3061 HVLTLRQALXKDSGTIYFEAGDQORASAAALRVTEKPSVPSRELTDTATITEGEDTLVLCBTS 3120
QY 3121 TCDIPMCWTGDKTILRGSAQCQLSHEGHRAQLLITGATLQDSGRYKCBAGGACSSSIVRV 3180
DB 3121 TCDIPMCWTGDKTILRGSAQCQLSHEGHRAQLLITGATLQDSGRYKCBAGGACSSSIVRV 3180
QY 3181 HARPVRFQEAUKDLEVELEGGAATLRCVLSVAAAPVKWCYGNVLRPGDKYSLRQEGAMLE 3240
DB 3181 HARPVRFQEAUKDLEVELEGGAATLRCVLSVAAAPVKWCYGNVLRPGDKYSLRQEGAMLE 3240
QY 3241 LVVRNLRPODSGRYSCSFGDQTTSATLTVPALPAQFICKLRNKEATEGATATLRCELSKT 3300
DB 3241 LVVRNLRPODSGRYSCSFGDQTTSATLTVPALPAQFICKLRNKEATEGATATLRCELSKT 3300
QY 3301 APVWRKGSSETLRDGDYCLRDQDQAMCELOIRGLAMVDAAEYSCVCGEERTSASITIRPM 3360
DB 3301 APVWRKGSSETLRDGDYCLRDQDQAMCELOIRGLAMVDAAEYSCVCGEERTSASITIRPM 3360
QY 3361 PAHFTIGRLRHQESIEGATATLRCELSKAAPVWRKGRSRLDGRHSLRQDGAVALCELOIC 3420
DB 3361 PAHFTIGRLRHQESIEGATATLRCELSKAAPVWRKGRSRLDGRHSLRQDGAVALCELOIC 3420
QY 3421 GLAVADAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCWELSKVAPVE 3480
DB 3421 GLAVADAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCWELSKVAPVE 3480
QY 3481 WRKGPENLRDGDYILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARF 3540
DB 3481 WRKGPENLRDGDYILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARF 3540
QY 3541 IEDVKNQDAREGATAVLQCELNSAAPVWRKGSSETLRDGDYSLRQDGTCKELQIRGLAM 3600
DB 3541 IEDVKNQDAREGATAVLQCELNSAAPVWRKGSSETLRDGDYSLRQDGTCKELQIRGLAM 3600
QY 3601 ADTGEYSCVCGQERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKVAPVWKG 3660
DB 3601 ADTGEYSCVCGQERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKVAPVWKG 3660
QY 3661 HETLRDGRHSLRQDQARCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKSFIEGL 3720
DB 3661 HETLRDGRHSLRQDQARCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKSFIEGL 3720
QY 3721 RNEBATEGDTATLWCELNSAAPVWRKGHETLRDGRHSLRQDGRSCELOIRGLAVVDAG 3780
DB 3721 RNEBATEGDTATLWCELNSAAPVWRKGHETLRDGRHSLRQDGRSCELOIRGLAVVDAG 3780
QY 3781 EYSCVCGQERTSATLTVPALPARFIEDVKNQDAREGATAVLQCELNSKAAPVWRKGSSETL 3840
DB 3781 EYSCVCGQERTSATLTVPALPARFIEDVKNQDAREGATAVLQCELNSKAAPVWRKGSSETL 3840
QY 3841 RGGDRYSLRQDGTTRCELOIHGSLVADTGEYSCVCGQERTSATLTVPAPQVPRPRELQSLQ 3900
DB 3841 RGGDRYSLRQDGTTRCELOIHGSLVADTGEYSCVCGQERTSATLTVPAPQVPRPRELQSLQ 3900
QY 3901 AEEGSTATLQCELSSEPTATVVMKGGLOQANGRRPRLOQCTABELVLQDLOREDTGEYT 3960
DB 3901 AEEGSTATLQCELSSEPTATVVMKGGLOQANGRRPRLOQCTABELVLQDLOREDTGEYT 3960
QY 3961 CTCGSOATSATLTVTAAAPVRLRELQHOVEDEGGTAHLCCELSRAGASVWRKGSLOLFP 4020
DB 3961 CTCGSOATSATLTVTAAAPVRLRELQHOVEDEGGTAHLCCELSRAGASVWRKGSLOLFP 4020
QY 4021 CAKQTMVQDGAALLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRLSQLEQ 4080
DB 4021 CAKQTMVQDGAALLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRLSQLEQ 4080
QY 4081 TGDILARLCCQLSDAESGAVVQWMLKEGVELHAGPKYEMRSQAGTARELLIHOLEAKDTGEYA 4140
DB 4081 TGDILARLCCQLSDAESGAVVQWMLKEGVELHAGPKYEMRSQAGTARELLIHOLEAKDTGEYA 4140
QY 4141 CVTGGQKTAASLRVTEPEVTIVRGLVDAEVTADEDEVEFSCEVSRAGATGVQWCLQGLPLQ 4200

Db 4141 CVTGGOKTAAASLRVTEPEVTI VRGLVDAEVTADEDEVEFSCEVSRAGATGVQWCLQGLPQ 4200
QY SNEVTEVAVRDGRIHTLRKLGVTPEADAGTVSHLGNHASSAOLTVRAPEVTTILEPLODVQ 4260
Db 4201 SNEVTEVAVRDGRIHTLRKLGVTPEADAGTVSHLGNHASSAOLTVRAPEVTTILEPLODVQ 4260
QY LSEGQDASFQCLRSRASGOEARWALGGVPLQANEMNDITVEQGTLLHLTLHKVTTLEDAGT 4320
Db 4261 LSEGQDASFQCLRSRASGOEARWALGGVPLQANEMNDITVEQGTLLHLTLHKVTTLEDAGT 4320
QY VSFHVGTCSSEAOQLKVTAKNTVVRGLENVAEALEGGEALFECOLSOPAEVAHHTWLLDDEPV 4380
Db 4321 VSFHVGTCSSEAOQLKVTAKNTVVRGLENVAEALEGGEALFECOLSOPAEVAHHTWLLDDEPV 4380
QY RTSENAENVFFENGLRHLLKLNLRPODSCRVTFLAGDMVTSAFITVRCWRLEIILEPLKN 4440
Db 4381 RTSENAENVFFENGLRHLLKLNLRPODSCRVTFLAGDMVTSAFITVRCWRLEIILEPLKN 4440
QY AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQPDSDWTVTADGSHOALLRSAQPHH 4500
Db 4441 AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQPDSDWTVTADGSHOALLRSAQPHH 4500
QY AGEVTFACBDASARLTVLGLPDPPEDAEVVAHSHSTVTLSWAAPMGSDGGGLCGYRVE 4560
Db 4501 AGEVTFACBDASARLTVLGLPDPPEDAEVVAHSHSTVTLSWAAPMGSDGGGLCGYRVE 4560
QY VKEGATGQWRLCHELVPGPECVVDGLAPEGTVFRVAAGVPVGAPEGPHLPQTVRLAEP 4620
Db 4561 VKEGATGQWRLCHELVPGPECVVDGLAPEGTVFRVAAGVPVGAPEGPHLPQTVRLAEP 4620
QY KVPVPPQAPESRQVAAGSDVLSLEVVAEAGEVIWHKGMERIQPGRFEVVSQGRQOML 4680
Db 4621 KVPVPPQAPESRQVAAGSDVLSLEVVAEAGEVIWHKGMERIQPGRFEVVSQGRQOML 4680
QY VTKGFTAEDQGVHCGLAGQSCIPAAATFOVALSPASVDEAPQPSLPPAEAQEGDLHLW 4740
Db 4681 VTKGFTAEDQGVHCGLAGQSCIPAAATFOVALSPASVDEAPQPSLPPAEAQEGDLHLW 4740
QY BALARKRMSREPTLDSISELPEEDGRSQRLPQEAEEVAPDLSEGYSTADELARTGDADL 4800
Db 4741 BALARKRMSREPTLDSISELPEEDGRSQRLPQEAEEVAPDLSEGYSTADELARTGDADL 4800
QY SHTSDDSDSRAGTSPSLVYLLKAGRPQTSPLASKVGAPAAPSVKPPQOOEPLAAVRPLG 4860
Db 4801 SHTSDDSDSRAGTSPSLVYLLKAGRPQTSPLASKVGAPAAPSVKPPQOOEPLAAVRPLG 4860
QY DLSTKDLGDPMSDKAAVKIQAAFGYKVRKEMKQOEGPMSHTFGDTAEAOVGDALRLCEV 4920
Db 4861 DLSTKDLGDPMSDKAAVKIQAAFGYKVRKEMKQOEGPMSHTFGDTAEAOVGDALRLCEV 4920
QY VASKADVRARWLKQGVELTDGRHHIHDOLGDGTCSLLIAGLDRADADAGCYTCQVSNKFGQV 4980
Db 4921 VASKADVRARWLKQGVELTDGRHHIHDOLGDGTCSLLIAGLDRADADAGCYTCQVSNKFGQV 4980
QY THSACVVVSGSESEASSGGELDDAFRAARRLHRLPRTKSPAESVDEELFLSADEGPA 5040
Db 4981 THSACVVVSGSESEASSGGELDDAFRAARRLHRLPRTKSPAESVDEELFLSADEGPA 5040
QY EPEPADQWTVREDEHFTICIRFEALTEARQAVTRFOEMFATLGI GVETKLVQEGPRVEM 5100
Db 5041 EPEPADQWTVREDEHFTICIRFEALTEARQAVTRFOEMFATLGI GVETKLVQEGPRVEM 5100
QY CISKETPAPVPPPEPLPSLLTSDAAPVFLTELQNEQVQDGYPVSDCVVTGQPMPSVRWF 5160
Db 5101 CISKETPAPVPPPEPLPSLLTSDAAPVFLTELQNEQVQDGYPVSDCVVTGQPMPSVRWF 5160
QY KQGLLEDDHYMINEOQGHQIITIAVVPADMGVVRCLAENMGVSSTKAEILRVDLTS 5220
Db 5161 KQGLLEDDHYMINEOQGHQIITIAVVPADMGVVRCLAENMGVSSTKAEILRVDLTS 5220
QY TDYDTAADATSSSYFSAQGYLSSREOGBTSTTDEGQLPOVBEELDLQVAPGTRLAKF 5280

Db 5221 TDYDTAADATSSSYFSAQGYLSSREOGBTSTTDEGQLPOVBEELDLQVAPGTRLAKF 5280
QY QLKVKGYPAPELXWFKQGOPLTASAHIRMTCKKILHTLEIISVTREDSGOYAAIISNMG 5340
Db 5281 QLKVKGYPAPELXWFKQGOPLTASAHIRMTCKKILHTLEIISVTREDSGOYAAIISNMG 5340
QY AAYSARILLVRGPDPEPEKPSADVHEOLVPPRMLERTFKPKVKKSSSTTFPSVKVEGRVVP 5400
Db 5341 AAYSARILLVRGPDPEPEKPSADVHEOLVPPRMLERTFKPKVKKSSSTTFPSVKVEGRVVP 5400
QY TVHMLREABEAGVLIWIGDTPGYTVASSAOQHSIIVLLDVGROHQGTCTCIASNAAGQALC 5460
Db 5401 TVHMLREABEAGVLIWIGDTPGYTVASSAOQHSIIVLLDVGROHQGTCTCIASNAAGQALC 5460
QY SASLHVSLPKVREOEKVKEALISTFLQGTTOAISAOGLETTASADIGGOKKEEPLAKE 5520
Db 5461 SASLHVSLPKVREOEKVKEALISTFLQGTTOAISAOGLETTASADIGGOKKEEPLAKE 5520
QY ALGHLSLAEVTEEFLOKLTSQITEMVSAKITQAKLOVPBGDSDEDSKTPSASPRHGRSR 5580
Db 5521 ALGHLSLAEVTEEFLOKLTSQITEMVSAKITQAKLOVPBGDSDEDSKTPSASPRHGRSR 5580
QY PSSSIQSSSESESEGDARGBEIFDIYVVTADYPLPGAEOQDAITLREGQYVEVLDAAHPLRW 5640
Db 5581 PSSSIQSSSESESEGDARGBEIFDIYVVTADYPLPGAEOQDAITLREGQYVEVLDAAHPLRW 5640
QY LVRTKPTKSSPSRQGWSPAYLDRRLKLSPEWGAAPPEPGEAVSEDEYKARLSSVIOE 5700
Db 5641 LVRTKPTKSSPSRQGWSPAYLDRRLKLSPEWGAAPPEPGEAVSEDEYKARLSSVIOE 5700
QY LLSSEQAFVELOQLQSHHLQHLERCPHPVPIAVAGQKAVIFRNVVDITGRPHSPFLOBLQ 5760
Db 5701 LLSSEQAFVELOQLQSHHLQHLERCPHPVPIAVAGQKAVIFRNVVDITGRPHSPFLOBLQ 5760
QY CDTDDDDVAMCFIKNOAAFEQYLFELVGRVQAESVWVSTAIQEFYKKYAEABALLAGDPSQP 5820
Db 5761 CDTDDDDVAMCFIKNOAAFEQYLFELVGRVQAESVWVSTAIQEFYKKYAEABALLAGDPSQP 5820
QY PPPLOHLYLQPPVERVORYOALLKELIRKARNRQNCALLEQAYAVVVSALLPQAEKHLV 5880
Db 5821 PPPLOHLYLQPPVERVORYOALLKELIRKARNRQNCALLEQAYAVVVSALLPQAEKHLV 5880
QY SLMENYPTGLLEALGEPITROGHFIWEGAPQARMPWKGNHRHVFILFRNHLVICPRDSRT 5940
Db 5881 SLMENYPTGLLEALGEPITROGHFIWEGAPQARMPWKGNHRHVFILFRNHLVICPRDSRT 5940
QY DTVSYVFRNMKLLSSIDLNDQVEGDDRAFRVWQEREDSVRKYLLOARTAIKSSWVKEIC 6000
Db 5941 DTVSYVFRNMKLLSSIDLNDQVEGDDRAFRVWQEREDSVRKYLLOARTAIKSSWVKEIC 6000
QY GIOQRLALPVWRPDPPEELADCTAELGETVKLACRVGTGPKPVIWYKDGKAVQVDPHH 6060
Db 6001 GIOQRLALPVWRPDPPEELADCTAELGETVKLACRVGTGPKPVIWYKDGKAVQVDPHH 6060
QY ILIEDPDGSCALILDSLTGVDSGOYMCFAASAAGNCSTLGKILVQVPPFRFVKNVSPFV 6120
Db 6061 ILIEDPDGSCALILDSLTGVDSGOYMCFAASAAGNCSTLGKILVQVPPFRFVKNVSPFV 6120
QY EGEDAQFTCTTIEGAPYPOIRWYKDGALLTTGNKFTLSEPRSGLLVIVIRAASKEDLGLY 6180
Db 6121 EGEDAQFTCTTIEGAPYPOIRWYKDGALLTTGNKFTLSEPRSGLLVIVIRAASKEDLGLY 6180
QY ECELVNRLGASARASAEIRIQSPMLQAOEQCHREOLVAEVDETTTLER - - - - -ADQEV 6232
Db 6181 ECELVNRLGASARASAEIRIQSPMLQAOEQCHREOLVAEVDETTTLER - - - - -ADQEV 6232
QY SVLKRLLLGPKAPGPPSTGDTLTPGCPGPRG - - - - -APAL - - - - -OETGSP - 6271
Db 6233 SVLKRLLLGPKAPGPPSTGDTLTPGCPGPRG - - - - -APAL - - - - -OETGSP - 6271
QY TVVKSPPGQRR - SPSKSPSRSPSRCSASPLRPGLLAPDLILYLPAGQPPRRPEAPGQKPV 6299
Db 6241 TVVKSPPGQRR - SPSKSPSRSPSRCSASPLRPGLLAPDLILYLPAGQPPRRPEAPGQKPV 6299
QY -PVTGSEAPAVPPRV - - - - -POPLLHGEPE - - - - -QEPEAIARAQEWTPVIRMEGA 6317
Db 6272 -PVTGSEAPAVPPRV - - - - -POPLLHGEPE - - - - -QEPEAIARAQEWTPVIRMEGA 6317
QY VFTLYVTEAEAHSPALPGLSGPQPKVVEEIEVVRKMGQGVSPITTE - - - - -VPRSSSGH 6357
Db 6300 VFTLYVTEAEAHSPALPGLSGPQPKVVEEIEVVRKMGQGVSPITTE - - - - -VPRSSSGH 6357

QY 6318 AW--PGACTG 6325
Db 6358 LFTLPGATPG 6367

RESULT 3

US-10-408-765A-2291
; Sequence 2291, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Faby, Boin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 66088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2291
; LENGTH: 6620
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2291

Query Match 77.9%; Score 32134; DB 16; Length 6620;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 6250; Conservative 15; Mismatches 57; Indels 48; Gaps 9;
QY 1 MDQFSGAPFLTRPKAFVSVGKDATLSQIVGNPTPQVSWKDDQOPVTAGARFLAQ 60
Db 1 MDQFSGAPFLTRPKAFVSVGKDATLSQIVGNPTPQVSWKDDQOPVTAGARFLAQ 60
QY 61 DGDYRLTLDLALGDSGOYVCRARNAIGEAFAAVGLQVDAEAAEQAHPFLRPTISIR 120
Db 61 DGDYRLTLDLALGDSGOYVCRARNAIGEAFAAVGLQVDAEAAEQAHPFLRPTISIR 120
QY 121 VREGSEATFRGVGSPRPVSWKDGRRLGEPDGPVRVVEELGASALIRRAAPRGG 180
Db 121 VREGSEATFRGVGSPRPVSWKDGRRLGEPDGPVRVVEELGASALIRRAAPRGG 180
QY 181 TYEVAENPLGAASAAALVDSDAADTASRPGTSTAALLAHLQRRREARAEAGAPSP 240
Db 181 TYEVAENPLGAASAAALVDSDAADTASRPGTSTAALLAHLQRRREARAEAGAPSP 240
QY 241 STGTRTCTVTGEGHARLSCTYVTGEPKPTVWKDQQLVTEGRRHVYVEDAQENFVLKILF 300
Db 241 STGTRTCTVTGEGHARLSCTYVTGEPKPTVWKDQQLVTEGRRHVYVEDAQENFVLKILF 300
QY 301 CKQSDRGILYCTASNLVQGTSSVLVVRPAVPPKQLDLVREKESATFLCEVPQPS 360
Db 301 CKQSDRGILYCTASNLVQGTSSVLVVRPAVPPKQLDLVREKESATFLCEVPQPS 360
QY 361 TEAAWFEETRLWASAKYIGIEEGTERLTVRNVSADDDAVYICETPEGSRTVAELAVQG 420
Db 361 TEAAWFEETRLWASAKYIGIEEGTERLTVRNVSADDDAVYICETPEGSRTVAELAVQG 420
QY 421 NLLRKLPRKTRAVRGDTAMFCVELAVPGVPVHWRNQEVEVAGGRVAISAGTHTLTIS 480
Db 421 NLLRKLPRKTRAVRGDTAMFCVELAVPGVPVHWRNQEVEVAGGRVAISAGTHTLTIS 480
QY 481 QCCLEDVQVAFMAGDCOTSTFCVSAAPRKPPLQPPDPVVKARMESSVILSWSPPHGE 540
Db 481 QCCLEDVQVAFMAGDCOTSTFCVSAAPRKPPLQPPDPVVKARMESSVILSWSPPHGE 540
QY 541 RPVTIDGYLVKSKKLGTYTWIRCHEAEWVATPELTADVABEGNFQPRVSAINSGQSPY 600
Db 541 RPVTIDGYLVKSKKLGTYTWIRCHEAEWVATPELTADVABEGNFQPRVSAINSGQSPY 600

Db 541 RPVTIDGYLVKSKKLGTYTWIRCHEAEWVATPELTADVABEGNFQPRVSAINSGQSPY 600
QY 601 LEFGPTVHLAPKLAVRTPLKAVQAVEGGEVTFPSVDLTVASAGFWLDOQALKASSVYETH 660
Db 601 LEFGPTVHLAPKLAVRTPLKAVQAVEGGEVTFPSVDLTVASAGFWLDOQALKASSVYETH 660
QY 661 CDRTRHTLTITREVPASLHGAOLKFWANGIESSIRNEVRAAPGLTANKPPAAAAREVLARL 720
Db 661 CDRTRHTLTITREVPASLHGAOLKFWANGIESSIRNEVRAAPGLTANKPPAAAAREVLARL 720
QY 721 HEEAQLLAELSDQAAAATVWLKDGRTLSPPKPYEQASAGRRVLLVRDVARDDAGLYECVS 780
Db 721 HEEAQLLAELSDQAAAATVWLKDGRTLSPPKPYEQASAGRRVLLVRDVARDDAGLYECVS 780
QY 781 RGGRIAYQLSVQGLARFLHDKMAGSCVDVAGGPAQFECETSEAHVHVHMYKDGMELGHS 840
Db 781 RGGRIAYQLSVQGLARFLHDKMAGSCVDVAGGPAQFECETSEAHVHVHMYKDGMELGHS 840
QY 841 GERFLOEDVGTTRHRLVAATVTRDEGTYSCRVGDSVDVFLRVSEPKVPFAKEQLARRKL 900
Db 841 GERFLOEDVGTTRHRLVAATVTRDEGTYSCRVGDSVDVFLRVSEPKVPFAKEQLARRKL 900
QY 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVKCMCATGCTRRLLVVOQAQADAGEY 960
Db 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVKCMCATGCTRRLLVVOQAQADAGEY 960
QY 961 SCEAGGQRLSPHLDVKEPKVFAKQVVAHSEVQAEAGANATLSCEVAQAQAEVMMYKDGK 1020
Db 961 SCEAGGQRLSPHLDVKEPKVFAKQVVAHSEVQAEAGANATLSCEVAQAQAEVMMYKDGK 1020
QY 1021 KLSSSLKVHVEAKGCRRLVVOQAQKTADAGDYSCBARGQVSRFLHITPEKMFKEQSV 1080
Db 1021 KLSSSLKVHVEAKGCRRLVVOQAQKTADAGDYSCBARGQVSRFLHITPEKMFKEQSV 1080
QY 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSSKVMEVKGCTRRLLVPOAGKAD 1140
Db 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSSKVMEVKGCTRRLLVPOAGKAD 1140
QY 1141 AGEYSCEAGGQVRSFHLHITPEKGVFAKEQSVHNEVQAEAGTAMLSCEVAQAQTEVTWY 1200
Db 1141 AGEYSCEAGGQVRSFHLHITPEKGVFAKEQSVHNEVQAEAGTAMLSCEVAQAQTEVTWY 1200
QY 1201 KDGKLSKSSSKVMEVKGCTRRLLVVOQVKGADAGEYSCEAGGQVRSFOLHITPEKAVPAK 1260
Db 1201 KDGKLSKSSSKVMEVKGCTRRLLVVOQVKGADAGEYSCEAGGQVRSFOLHITPEKAVPAK 1260
QY 1261 EQLVHNEVTRTAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAGCMRQLVVOQA 1320
Db 1261 EQLVHNEVTRTAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAGCMRQLVVOQA 1320
QY 1321 QOADAGEYTCBAGGQRLSPHLDVSEPKAVFAKEQLAHRKVQAEAGAIATLSCEVAQAQTE 1380
Db 1321 QOADAGEYTCBAGGQRLSPHLDVSEPKAVFAKEQLAHRKVQAEAGAIATLSCEVAQAQTE 1380
QY 1381 VTWYKDGKLSKSSSKVMEVKGCTRRLLVVOQAQADAGEYSCEAGGQRLSPSLDVAEPKV 1440
Db 1381 VTWYKDGKLSKSSSKVMEVKGCTRRLLVVOQAQADAGEYSCEAGGQRLSPSLDVAEPKV 1440
QY 1441 VPAKEQPVHREVOQAQAGSTTTLSCVEAQAQTEVTWYKDGKLSKSSSKVMEVKGCTRRLLV 1500
Db 1441 VPAKEQPVHREVOQAQAGSTTTLSCVEAQAQTEVTWYKDGKLSKSSSKVMEVKGCTRRLLV 1500
QY 1501 VQQAQADAGEYSCEAGGQRLSPHLDVSEPKAVFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQV 1560
Db 1501 VQQAQADAGEYSCEAGGQRLSPHLDVSEPKAVFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQV 1560
QY 1561 AQTETWYKDGKLSKSSSKVMEVKGCTRRLLVVOQAQADAGEYSCKAGDQRLSPHLDVA 1620
Db 1561 AQTETWYKDGKLSKSSSKVMEVKGCTRRLLVVOQAQADAGEYSCKAGDQRLSPHLDVA 1620
QY 1621 EPKVVFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQV 1680
Db 1621 EPKVVFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQVPAKFAKEQV 1680

Qy	1681	RRLVVQAQADAGEYCEAGGQRLSRLHVAELEPOISRRPCRRPLVVKHEHDIILTA	1740	2761	HESASDPAMWTVGKTVGSSSRFOATROGKRYKILVVRREAAPSDAGEVVFVSRGLTSKASL	2820
Db	1681		1740	2761		2820
Qy	1741	TLATPSAATVTLKDGVEIRRSKHETASQGDTHTLTVHGAQVLDLSAIYSRCVGAEGQDF	1800	2821	IYVERPAAIIKPIEDQWVARGEDVELRCELSRAGTPVHMLKDRKAIKRSOKYDVVCEGTM	2880
Db	1741		1800	2821		2880
Qy	1801	PVQVEEVAAPCRLLPEVGBLGGTVTLACBLSPACAEVVMRCGNTOPRVKRFQMWAE	1860	2881	AMLVIRGASLKDAGEYTCVEASKSTASLHVEBKANCFTBELTNLQVEEKGTVFTCKTE	2940
Db	1801		1860	2881		2940
Qy	1861	PVRSLTVLGLRAEDAGEVVCESRDHDSAQITVSVPRVVKFMSGISTVVAEEGGEATFOC	1920	2941	HPAATVTVWRKGLLELRASGHQPSQEGITLRLITISALEKADSDTYTCDIQOASRAQLLV	3000
Db	1861		1920	2941		3000
Qy	1921	VVSPSDVAVWFRDGAIIQPSKEFAISQSGASHSLTISDLVLEDAGQITVEABEGASSAA	1980	3001	QGRVHIIEDLEDDVDVOEGSSATFRCRISPANYPEVHWFDKTPLHANELNEIDAQPGGY	3060
Db	1921		1980	3001		3060
Qy	1981	LRVREAPVLFKKLEPQTVBERSSVTLELTPWPELRWTRNATALAPGKNVEIHAEGA	2040	3061	HVLTLRQALAKDSGTIYFEAGDQASAAALRVTEKPSVFSRELTDATITTEGEDTLVCE	3120
Db	1981		2040	3061		3120
Qy	2041	RHRLVHNHVGADRGFFCETPDDKTQAKLTVMRQVRLVRGLQAVEAREOGTATMEVQL	2100	3121	TCDIPMCWTGDKTILRGASARCOLSHEGHRAQLLITGATLODSGRYKCEAGCACSSIVRV	3180
Db	2041		2100	3121		3180
Qy	2101	SHADVDGSWTRDGLRFQOGPTCHLAVRGPMTLTLISGLRPEDSGLMVFKABGVHTSARLV	2160	3241	LVVNLRLPQDSGRYSCSFGDQTTSATLTVTALPAQFTGKLRNKEATEGATATLCELSKT	3300
Db	2101		2160	3241		3300
Qy	2161	VTELTVSRSRPLQDVVTTKEKVTLECELSPNDVVRMLKDGVELRAGKTWALIAAQACR	2220	3301	APVWRKGSSETLRDGDYCLRDQAMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPM	3360
Db	2161		2220	3301		3360
Qy	2221	SITIYRCFPADQGVVCDADHAQSSAVKVOGRTYTLIYRRLVLAEDAGEIOFVAENAESR	2280	3361	PAHFIRGLRHQESIEGATATLRCBLSKAAAPVEMWKGRSLRDGDRHSLRDQGVACELQIC	3420
Db	2221		2280	3361		3420
Qy	2281	AQLRVKELPVTILVRPLRKIAKHEKRGVLECOVSRASAOVRWFKGSQBLQPGPKYELVSD	2340	3421	GLAVADAGEYSCVCGEERTSATLTVKALPAKFTGELRNEEAVEGATAMLWCELSKVA	3480
Db	2281		2340	3421		3480
Qy	2341	GLYRKLIIISDVHAEDEDTYCDAGDVKTSAQPFVEEQSITIVRGLODVTVMPEPAPWPEC	2400	3481	WRKGPENLRDGDYILRQEGTRCELOICGLAMADAGEYLCVCGOERTSATLITRALPARF	3540
Db	2341		2400	3481		3540
Qy	2401	ETSIPIVPPKWLKGTVLQAGNVGLEQEGTVHRLMLRRTCTMTGPHVFTVGKSRSSA	2460	3541	IEDVKNQREAREGATAVLCCELNSAAPVEMWKGSSETLRDGDYSLRDQGTCKELOIRGLAM	3600
Db	2401		2460	3541		3600
Qy	2461	RLVWSDIPVLTREPLEKTRGELQSVLSCDPRPAPKAVQWYKDDTPLSPSEKFKWLSLEG	2520	3601	ADTGEYSCVCGOERTSAMLTVRALPIKFTGELRNEEATEGATVLRCELKMAWPVEMWK	3660
Db	2461		2520	3601		3660
Qy	2521	QMAELRIILMPADAGVYRCQAGSAHSSTEVTVAREVTVTGLODAEATEEGWASFCE	2580	3661	HETLRDGDHSLRDQDARGARCELOIRGLVAEDAGEYLCMCCKERTSAMLTVRAMPKFI	3720
Db	2521		2580	3661		3720
Qy	2581	LSHEDEEVEWSLNGMPLYNDSFHEISHKGRRHITLVKSTORADAGIVRASSILKVSTARS	2640	3721	RNEEATEGDTATLWCELSKAAAPVEMWKGHETLRDGDHSLRDQDARGARCELOIRGLVA	3780
Db	2581		2640	3721		3780
Qy	2641	FVRVFPVFLKALDDLAEERGTALQCEVSPBEAHVVRKDGVOLGPSDKDYDLHTAGT	2700	3781	EYSCVCGOERTSATLTVRALPARFIEDVKNQREAREGATAVLCCELSKAAPVEMWKGS	3840
Db	2641		2700	3781		3840
Qy	2701	RGLVHVDVSPEDAGLYTCHVGSEETRARVRVHDLHVGITKRLKTMTEVLESGSCS	2760	3841	RGGDRYSLRDQGTCELOIRGLHSLVADTGEYSCVCGOERTSATLTVRAPQVFPVFP	3900
Db	2701		2760	3841		3900

||||| 3841 RGGDRYSLRQDTRLEQLHGLSVADTGBEYSCVCCQERTSALTIVRAPQVPRFBPLQSLQ 3900
QY ||||| 3901 ABEESTATLQCELSBPTATVVMWSKGLQLQANGREPRILQCGTABELVLQDLQREDTGEYT 3960
Db ||||| 3901 ABEESTATLQCELSBPTATVVMWSKGLQLQANGREPRILQCGTABELVLQDLQREDTGEYT 3960
QY ||||| 3961 CTCGSOATSATLTVTAAPVRFLELQHQVEVDGGTAHLCCELSRAGASVEMWKGSLQLP 4020
Db ||||| 3961 CTCGSOATSATLTVTAAPVRFLELQHQVEVDGGTAHLCCELSRAGASVEMWKGSLQLP 4020
QY ||||| 4021 CAKYQMVQDGAABELVRGVEQEDAGDYTCDTGHTQSMASLSVRPRPKFKTRLQSLQEOE 4080
Db ||||| 4021 CAKYQMVQDGAABELVRGVEQEDAGDYTCDTGHTQSMASLSVRPRPKFKTRLQSLQEOE 4080
QY ||||| 4081 TGDJARLCCQLSDAESGAVQWMLKEGVELHAGPKYEMRSQATRELLIHQLEAKDTGEYA 4140
Db ||||| 4081 TGDJARLCCQLSDAESGAVQWMLKEGVELHAGPKYEMRSQATRELLIHQLEAKDTGEYA 4140
QY ||||| 4141 CVTGGOKTAASLRVTEPEVTIVRGLVDAEVTADEVFECEVSRAGATGVQWCLQGLPLQ 4200
Db ||||| 4141 CVTGGOKTAASLRVTEPEVTIVRGLVDAEVTADEVFECEVSRAGATGVQWCLQGLPLQ 4200
QY ||||| 4201 SNEVTEAVRQGRITHLRLKGVTPEDAGTVSPHLGNHASSAQLTVRAPEVTILEPLQDVQ 4260
Db ||||| 4201 SNEVTEAVRQGRITHLRLKGVTPEDAGTVSPHLGNHASSAQLTVRAPEVTILEPLQDVQ 4260
QY ||||| 4261 LSEGQDASFQCRLSRASQOERWALGGVPLQANEMNDITVEQGTLLHLTLHKVTLDEAGT 4320
Db ||||| 4261 LSEGQDASFQCRLSRASQOERWALGGVPLQANEMNDITVEQGTLLHLTLHKVTLDEAGT 4320
QY ||||| 4321 VSFHVGTCSSEAOQLKVTAKNTVVRGLENVAEALGEGEALFECQLSQPEVAHAHTWLLDDEPV 4380
Db ||||| 4321 VSFHVGTCSSEAOQLKVTAKNTVVRGLENVAEALGEGEALFECQLSQPEVAHAHTWLLDDEPV 4380
QY ||||| 4381 RTSENAEVVFPENGRLHLLLNLPQDSRCVTFFLAGDMVTSAFITVRGWRLEIILEPLKN 4440
Db ||||| 4381 RTSENAEVVFPENGRLHLLLNLPQDSRCVTFFLAGDMVTSAFITVRGWRLEIILEPLKN 4440
QY ||||| 4441 AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQDDSDWTVTADGSHQALLRSQAHPH 4500
Db ||||| 4441 AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQDDSDWTVTADGSHQALLRSQAHPH 4500
QY ||||| 4501 AGEVTFACRDVAASARLTVLGLPDPPEDAEVVAHSSHTVTLSWAAPMSDGGGLCGYRVE 4560
Db ||||| 4501 AGEVTFACRDVAASARLTVLGLPDPPEDAEVVAHSSHTVTLSWAAPMSDGGGLCGYRVE 4560
QY ||||| 4561 VKEGATGWRLCHELVPGPECVVDGLAPGETYRFRVAAGVPVGAEPVHLPTOTVRLABPP 4620
Db ||||| 4561 VKEGATGWRLCHELVPGPECVVDGLAPGETYRFRVAAGVPVGAEPVHLPTOTVRLABPP 4620
QY ||||| 4621 KPVPQPSAPESRQVAAGEDVSLELVVAEAGEVIVHKGMERIQPGGRFVVVSQGRQWML 4680
Db ||||| 4621 KPVPQPSAPESRQVAAGEDVSLELVVAEAGEVIVHKGMERIQPGGRFVVVSQGRQWML 4680
QY ||||| 4681 VIKGTAEDQGEYHCGLAQGSTCPAAATFQVALSPASVDEABQPSLPPEAAQEGDLHLW 4740
Db ||||| 4681 VIKGTAEDQGEYHCGLAQGSTCPAAATFQVALSPASVDEABQPSLPPEAAQEGDLHLW 4740
QY ||||| 4741 EALAKRRMSREPTLDSISELPEEDGRSQRQLPQEAEEVAPDLSEGYSTADELARTGDADL 4800
Db ||||| 4741 EALAKRRMSREPTLDSISELPEEDGRSQRQLPQEAEEVAPDLSEGYSTADELARTGDADL 4800
QY ||||| 4801 SHTSDDESRACTPSLVLYLKKAGRPGTSPPLASKVGAAPAAKPSKQOQOEPLAAVRPPLG 4860
Db ||||| 4801 SHTSDDESRACTPSLVLYLKKAGRPGTSPPLASKVGAAPAAKPSKQOQOEPLAAVRPPLG 4860
QY ||||| 4861 DLSTKDLGDPMSDKAAVKIQAAFKYKVRKEMKQOEGPMFSHTFGDTBAQVGDALRLCEV 4920
Db ||||| 4861 DLSTKDLGDPMSDKAAVKIQAAFKYKVRKEMKQOEGPMFSHTFGDTBAQVGDALRLCEV 4920
QY ||||| 4921 VASKADVRARWLKDGVELTDRGHHHI DQLDGDTCSLLIAGLDRADAGCYTCOVSNKFGQV 4980
|||||

Db ||||| 4921 VASKADVRARWLKDGVELTDRGHHHI DQLDGDTCSLLIAGLDRADAGCYTCOVSNKFGQV 4980
QY ||||| 4981 THSACVVVSGSESEAEBSGSGELDDAFRAARRLHRLFRTKSPAEVSEBELFLSADGPA 5040
Db ||||| 4981 THSACVVVSGSESEAEBSGSGELDDAFRAARRLHRLFRTKSPAEVSEBELFLSADGPA 5040
QY ||||| 5041 EPEEPADHOTVREDEHFCIRFEALTEARQAVTRFEQEMPATLIGIVEIKLVEQGGPRVRM 5100
Db ||||| 5041 EPEEPADHOTVREDEHFCIRFEALTEARQAVTRFEQEMPATLIGIVEIKLVEQGGPRVRM 5100
QY ||||| 5101 CISKETPAPVVPPEPLPSLLTSDAAPFLTELQNOVEQDGVPSVPCVVVTGQPMFSVRWF 5160
Db ||||| 5101 CISKETPAPVVPPEPLPSLLTSDAAPFLTELQNOVEQDGVPSVPCVVVTGQPMFSVRWF 5160
QY ||||| 5161 KDKLLBEDDHYMINEDQGGHQLIITAVVPADMGVYRCLAEBSMGVSTKAEALRVDLTS 5220
Db ||||| 5161 KDKLLBEDDHYMINEDQGGHQLIITAVVPADMGVYRCLAEBSMGVSTKAEALRVDLTS 5220
QY ||||| 5221 TDYDTAADAATSSSYFSAQGYLSSREQGTSTTDEGOLPOVVEELRDLQVAPGTRLAKF 5280
Db ||||| 5221 TDYDTAADAATSSSYFSAQGYLSSREQGTSTTDEGOLPOVVEELRDLQVAPGTRLAKF 5280
QY ||||| 5281 QLKVKGYPAPLYWPKDQPLTASAHIRMTGKKILHTLEIISVTREDSGQAAAYISNANG 5340
Db ||||| 5281 QLKVKGYPAPLYWPKDQPLTASAHIRMTGKKILHTLEIISVTREDSGQAAAYISNANG 5340
QY ||||| 5341 AAYSARLLVRGPDPPEPKPQSDVHEQLVPPRMLERFTPKKVKKGSSITFSVKVGRPVP 5400
Db ||||| 5341 AAYSARLLVRGPDPPEPKPQSDVHEQLVPPRMLERFTPKKVKKGSSITFSVKVGRPVP 5400
QY ||||| 5401 TVHMLREAEAGVLMWIGDPTGYTVASSAQOHSVLVLDVGRHQGTYTCTIASNAGAOLC 5460
Db ||||| 5401 TVHMLREAEAGVLMWIGDPTGYTVASSAQOHSVLVLDVGRHQGTYTCTIASNAGAOLC 5460
QY ||||| 5461 SASLHVSGLPKVBEQEKVKEALISTFLQGTTOAISAOGLTASPADLGGORKEEPLAAKE 5520
Db ||||| 5461 SASLHVSGLPKVBEQEKVKEALISTFLQGTTOAISAOGLTASPADLGGORKEEPLAAKE 5520
QY ||||| 5521 ALGHLSLAEVGTBEFLQKLTQITQTEMVSAKITQAKLQVPGGSDSDSTPSAPRGRSR 5580
Db ||||| 5521 ALGHLSLAEVGTBEFLQKLTQITQTEMVSAKITQAKLQVPGGSDSDSTPSAPRGRSR 5580
QY ||||| 5581 PSSSTQESSSESDGDARGEIIFDIYVVTADYLPJGAEQDAITLREGQVVEVLDAAHPLRW 5640
Db ||||| 5581 PSSSTQESSSESDGDARGEIIFDIYVVTADYLPJGAEQDAITLREGQVVEVLDAAHPLRW 5640
QY ||||| 5641 LVRTKPTKSSPSRQGWSPAYLDRRLKLSPEWGAEEAEFFPGEAEVSEDEYKARLSSVIOE 5700
Db ||||| 5641 LVRTKPTKSSPSRQGWSPAYLDRRLKLSPEWGAEEAEFFPGEAEVSEDEYKARLSSVIOE 5700
QY ||||| 5701 LLSSEQAFVBEFLQFLOSHHQLHLERCPHVPIAVAGQKAVIPRNVRDIGRPHSSFLQELQO 5760
Db ||||| 5701 LLSSEQAFVBEFLQFLOSHHQLHLERCPHVPIAVAGQKAVIPRNVRDIGRPHSSFLQELQO 5760
QY ||||| 5761 CDTDDVAMCTKNQAAFEQYLEFLVGRVQAESVVVSTAIQEFYKKAEEALLAGDPSQ 5820
Db ||||| 5761 CDTDDVAMCTKNQAAFEQYLEFLVGRVQAESVVVSTAIQEFYKKAEEALLAGDPSQ 5820
QY ||||| 5821 PPPPLQHYLEQPVVERVQRYQALLKELIRNKARNRONCALLQEAQAVVVSALPORASNKLVH 5880
Db ||||| 5821 PPPPLQHYLEQPVVERVQRYQALLKELIRNKARNRONCALLQEAQAVVVSALPORASNKLVH 5880
QY ||||| 5881 SLMENYPGTLBALGEPTRQGHFIWVWEGAPGARMKGNHVRVFLPRNHLVICKPRRDSRT 5940
Db ||||| 5881 SLMENYPGTLBALGEPTRQGHFIWVWEGAPGARMKGNHVRVFLPRNHLVICKPRRDSRT 5940
QY ||||| 5941 DTVSYVFRNMKLSIDLNDQVEGDDRAFEVWQEREDSVRKYLLOARTAIKSSWVKETC 6000
Db ||||| 5941 DTVSYVFRNMKLSIDLNDQVEGDDRAFEVWQEREDSVRKYLLOARTAIKSSWVKETC 6000
QY ||||| 6001 GIORLALPVWRPDPFBEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDPHH 6060
Db ||||| 6001 GIORLALPVWRPDPFBEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDPHH 6060

QY 6061 ILIEDPDGSCALILDSLTGVDGQYMCFAASAAGNCSTLGKILVQVPPRFVNVKVRASPV 6120
DB 6061 ILIEDPDGSCALILDSLTGVDGQYMCFAASAAGNCSTLGKILVQVPPRFVNVKVRASPV 6120
QY 6121 EGEDAQFTCTTGAPYQIRWYKDGALLTGKFKQTLSEPRSGLLVIVIRAAKEDLGLY 6180
DB 6121 EGEDAQFTCTTGAPYQIRWYKDGALLTGKFKQTLSEPRSGLLVIVIRAAKEDLGLY 6180
QY 6181 ECENLNLGSARASABLRISQPMLOAQEOCHREQLVAIVEDTTLR-----ADQEV 6232
DB 6181 ECENLNLGSARASABLRISQPMLOAQEOCHREQLVAIVEDTTLR-----ADQEV 6240
QY 6233 SVLKRLGPKAPGPTGDLTGPCPCPRG-----APAL-----QETGSP- 6271
DB 6241 TVVKSPPRGQR-SPSKSPSPSRCSASPLRPGLLAPDLLYLPAGAGPRRPEAEPGQKPV 6299
QY 6272 -PVTGTSEAPAVPRV-----PQPLLHEGPE-----QPEATARAQEWTPVTRMGA 6317
DB 6300 VPTLYVTEAEAHSPALGSLGSPQPKWVEETIEYVRVKMGQGVSPTE--VPRSSSGH 6357
QY 6318 AW--PGAGTG 6325
DB 6358 LFTLPGATPG 6367

RESULT 4

US-10-077-130-2
; Sequence 2, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: MPI2001-047PJRCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2630
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-2

Query Match 33.2%; Score 13710; DB 13; Length 2630;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5339 MGAAYSSARLLVRGDEPEEKPSADVHEQLVPPRMLEPFTPKVKKGSSITFSVKVEGRP 5398
DB 1 MGAAYSSARLLVRGDEPEEKPSADVHEQLVPPRMLEPFTPKVKKGSSITFSVKVEGRP 60
QY 5399 VPTVHMLREEAERGVLWTPGDTPGYTVASSAQHSLVLLDVGROHQGTTCIASNAGQA 5458
DB 61 VPTVHMLREEAERGVLWTPGDTPGYTVASSAQHSLVLLDVGROHQGTTCIASNAGQA 120
QY 5459 LCSASLHVSGLPKVEEQEKVEALISTELQGTQAI SAQGLTASPADLGGQKKEEPLAA 5518
DB 121 LCSASLHVSGLPKVEEQEKVEALISTELQGTQAI SAQGLTASPADLGGQKKEEPLAA 180
QY 5519 KEALGHLSAEVGTGEFLQKLTQITTEMVSAKITQAKLVPGGSDSDSKTPSPRHR 5578
DB 181 KEALGHLSAEVGTGEFLQKLTQITTEMVSAKITQAKLVPGGSDSDSKTPSPRHR 240
QY 5579 SRPSSIOESSSESDGDARGEIIDIYVVTADYLPGLAEQDAITLRGQYVEVLDAAHPL 5638
DB 241 SRPSSIOESSSESDGDARGEIIDIYVVTADYLPGLAEQDAITLRGQYVEVLDAAHPL 300
QY 5639 RNLVRTKTPSKSPRQGVSPAYLDRRLKLSPEWGAEEAFPEGAEVSEDEYKARLSSVI 5698

DB 301 RNLVRTKTPSKSPRQGVSPAYLDRRLKLSPEWGAEEAFPEGAEVSEDEYKARLSSVI 360
QY 5699 QELLSSEQAFVEELQFLOSHLQHLERCPhVPIAVAGQKAVI FRNVRDIGHPSFSFLOEL 5758
DB 361 QELLSSEQAFVEELQFLOSHLQHLERCPhVPIAVAGQKAVI FRNVRDIGHPSFSFLOEL 420
QY 5759 QQCOTDDDVAMCFIKNOAFAEQYLEFLVGRVQAESVVVSTAIQBFYKYKAEALLAGDPS 5818
DB 421 QQCOTDDDVAMCFIKNOAFAEQYLEFLVGRVQAESVVVSTAIQBFYKYKAEALLAGDPS 480
QY 5819 QPPPPPLQHYLEQPVVERVQYQALLKELIRNKARNRQNCALLEQAYAVVSALPORAENKL 5878
DB 481 QPPPPPLQHYLEQPVVERVQYQALLKELIRNKARNRQNCALLEQAYAVVSALPORAENKL 540
QY 5879 HVSILMENYPGTLEALGSPIRQGHFI VMGAPGAPMKGHNRHVFLFRNHLVICKPRDS 5938
DB 541 HVSILMENYPGTLEALGSPIRQGHFI VMGAPGAPMKGHNRHVFLFRNHLVICKPRDS 600
QY 5939 RTDVTSVVFRNMKLSISIDLDNDQVEGDRAFEVWQEREDSVRKYLLOARTAIKSSWVKE 5998
DB 601 RTDVTSVVFRNMKLSISIDLDNDQVEGDRAFEVWQEREDSVRKYLLOARTAIKSSWVKE 660
QY 5999 ICGIQORLALPVMRPPDFEBELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDP 6058
DB 661 ICGIQORLALPVMRPPDFEBELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDP 720
QY 6059 HHILIEDPDGSCALILDSLTGVDGQYMCFAASAAGNCSTLGKILVQVPPRFVNVKVRASPV 6118
DB 721 HHILIEDPDGSCALILDSLTGVDGQYMCFAASAAGNCSTLGKILVQVPPRFVNVKVRASPV 780
QY 6119 FVEGEDAQFTCTTGAPYQIRWYKDGALLTGKFKQTLSEPRSGLLVIVIRAAKEDLGL 6178
DB 781 FVEGEDAQFTCTTGAPYQIRWYKDGALLTGKFKQTLSEPRSGLLVIVIRAAKEDLGL 840
QY 6179 LYECELVNLGLSARASABLRISQPMLOAQEOCHREQLVAIVEDTTLRADQEVTSVLKRL 6238
DB 841 LYECELVNLGLSARASABLRISQPMLOAQEOCHREQLVAIVEDTTLRADQEVTSVLKRL 900
QY 6239 LGPKAPGSTGDLTGPCPCPRGAPALQETGSPVVTGTSAPAVPPRPVPOPLLHEGPEQE 6298
DB 901 LGPKAPGSTGDLTGPCPCPRGAPALQETGSPVVTGTSAPAVPPRPVPOPLLHEGPEQE 960
QY 6299 PEAIARAQEWTPVIRMEGAAMPGAGTGELLWDVHSHVRETTRQTYTQAI DTHTRAPPS 6358
DB 961 PEAIARAQEWTPVIRMEGAAMPGAGTGELLWDVHSHVRETTRQTYTQAI DTHTRAPPS 1020
QY 6359 MQVTIEDVQAOTGGTAQFEAIIEGDPQPSVTWYKDSVOLVDSTRLSQOQEGTYSVLVRH 6418
DB 1021 MQVTIEDVQAOTGGTAQFEAIIEGDPQPSVTWYKDSVOLVDSTRLSQOQEGTYSVLVRH 1080
QY 6419 VASKDAGVYTCLAQNTGGQVLCKAELLVLGDDNEPDESEKSHRRKLSFYEVEKEEIGRGV 6478
DB 1081 VASKDAGVYTCLAQNTGGQVLCKAELLVLGDDNEPDESEKSHRRKLSFYEVEKEEIGRGV 1140
QY 6479 FGFVKRVQHKGNKILCAAKFIPLRSRTRAQAYRERDILAAALSHPLVTGLLDQFETRKTLI 6538
DB 1141 FGFVKRVQHKGNKILCAAKFIPLRSRTRAQAYRERDILAAALSHPLVTGLLDQFETRKTLI 1200
QY 6539 LILELCSSEBELLDRLYRKGVVVTEAEVKVYIQQLVEGLHYLHSHGVHLHDIKPSNLMVHP 6598
DB 1201 LILELCSSEBELLDRLYRKGVVVTEAEVKVYIQQLVEGLHYLHSHGVHLHDIKPSNLMVHP 1260
QY 6599 AREDIKICDPGFAQNTTPAELQPSQSGSPFVSPEIIQQNPVSEASDIWAMGVISYLSLT 6658
DB 1261 AREDIKICDPGFAQNTTPAELQPSQSGSPFVSPEIIQQNPVSEASDIWAMGVISYLSLT 1320
QY 6659 CSSPFFAGESDRATLLNVLEGRVSNWSPMAHLSEDAKDFIKATLQAPAPRPSAAQCLSH 6718
DB 1321 CSSPFFAGESDRATLLNVLEGRVSNWSPMAHLSEDAKDFIKATLQAPAPRPSAAQCLSH 1380
QY 6719 PWFLLKMPAEAAHFINTKOLKFLLRSRWORSIMYSKSIILVMRSIPPELLRPPDSPSLGV 6778

QY	5913	MPWKGNHRHVLF	FRNHLVI	CKPRDRSR	TDTVSYVFRNMMK	SSIDLNDQ	VEGDDRAPEVW	5972						
Db	541	MPWKGNHRVLF	FRNHLVI	CKPRDRSR	TDTVSYVFRNMMK	SSIDLNDQ	VEGDDRAPEVW	600						
QY	5973	QEREDSVKYLLO	ARTAI	IKSSWKE	ICGIIQORLAL	PVMRPDP	FEELADCTAELGBTVK	6032						
Db	601	QEREDSVKYLLO	ARTAI	IKSSWKE	ICGIIQORLAL	PVMRPDP	FEELADCTAELGBTVK	660						
QY	6033	LACRVGTGPKP	VI	SWYKDGKAVQ	DPHHILIED	PGSGCAL	ILDSLTGVD	SOYMCFAASA	6092					
Db	661	LACRVGTGPKP	VI	SWYKDGKAVQ	DPHHILIED	PGSGCAL	ILDSLTGVD	SOYMCFAASA	720					
QY	6093	AGNCSTLGI	L	VQVPPRVN	KVRASPF	VEGEDA	QFTCTIEGAP	YQIRWKDGALLTGN	6152					
Db	721	AGNCSTLGI	L	VQVPPRVN	KVRASPF	VEGEDA	QFTCTIEGAP	YQIRWKDGALLTGN	780					
QY	6153	KFQTLSEPR	SGLLV	IRAASKE	DGLGYCECLVN	RLGSARASAE	LR	IOSPMLQAOEOCHR	6212					
Db	781	KFQTLSEPR	SGLLV	IRAASKE	DGLGYCECLVN	RLGSARASAE	LR	IOSPMLQAOEOCHR	840					
QY	6213	EOLVAAVED	TT	LERADQ	EVTSVLKRLG	PKAPGST	GDLTGFG	PCPRGAPALQETGSQPP	6272					
Db	841	EOLVAAVED	TT	LERADQ	EVTSVLKRLG	PKAPGST	GDLTGFG	PCPRGAPALQETGSQPP	900					
QY	6273	VTGTSEAP	AVP	PRVPOPLL	HGPGQEP	EAIARAQ	EWTPVIR	MEGAAPWGAGTGELLMDVH	6332					
Db	901	VTGTSEAP	AVP	PRVPOPLL	HGPGQEP	EAIARAQ	EWTPVIR	MEGAAPWGAGTGELLMDVH	960					
QY	6333	SHVVRETT	QRTY	TQADITH	TARPSSQ	MVTTIED	VQAO	TGGTGAQPEALIEGDP	POPSVTWYK	6392				
Db	961	SHVVRETT	QRTY	TQADITH	TARPSSQ	MVTTIED	VQAO	TGGTGAQPEALIEGDP	POPSVTWYK	1020				
QY	6393	DSVQLV	DS	TRLSQQ	QEGTYS	SLVLRHV	ASKDAGY	YTCLAQNTG	QQVLCKAELLVLGGDNE	6452				
Db	1021	DSVQLV	DS	TRLSQQ	QEGTYS	SLVLRHV	ASKDAGY	YTCLAQNTG	QQVLCKAELLVLGGDNE	1080				
QY	6453	PDSEKQSH	RK	LKHSFY	EKVEE	IGRGVGF	VKRVQ	HGNKILCAAK	FPLRSRTRAQAYRE	6512				
Db	1081	PDSEKQSH	RK	LKHSFY	EKVEE	IGRGVGF	VKRVQ	HGNKILCAAK	FPLRSRTRAQAYRE	1140				
QY	6513	RDILAA	L	SHPLVT	GLLOQ	FETRKT	LIILILEL	CSSEELDL	RYKGVVTEAEVKVYIQOLV	6572				
Db	1141	RDILAA	L	SHPLVT	GLLOQ	FETRKT	LIILILEL	CSSEELDL	RYKGVVTEAEVKVYIQOLV	1200				
QY	6573	EGLHYL	SH	GVHLHD	IKPSN	ILMVHP	AREDIK	ICDFGA	QNTITPAELQFSQYGSPEFVSP	6632				
Db	1201	EGLHYL	SH	GVHLHD	IKPSN	ILMVHP	AREDIK	ICDFGA	QNTITPAELQFSQYGSPEFVSP	1260				
QY	6633	EIIIOON	PV	SEASD	I	WANGV	ISYLS	LTCS	SPAGESDRATILNVLEGRVSWSSPMAHLSE	6692				
Db	1261	EIIIOON	PV	SEASD	I	WANGV	ISYLS	LTCS	SPAGESDRATILNVLEGRVSWSSPMAHLSE	1320				
QY	6693	DAKDFI	KAT	LQARA	QAP	APSAQ	CLSH	PWFLK	SMPEAEAHFINTKOLKFL	LARSRWORS	6752			
Db	1321	DAKDFI	KAT	LQARA	QAP	APSAQ	CLSH	PWFLK	SMPEAEAHFINTKOLKFL	LARSRWORS	1380			
QY	6753	SYKILV	MS	RIPELL	R	GPDP	SPSLGV	VARHL	CRDTGGSSSSSSSDNELAP	FAPAKSLPPS	6812			
Db	1381	SYKILV	MS	RIPELL	R	GPDP	SPSLGV	VARHL	CRDTGGSSSSSSSDNELAP	FAPAKSLPPS	1440			
QY	6813	PVTHS	PL	LH	PRGFL	R	PSASL	PEEA	EASERSTEAP	PASPEGAGPPAAOCVPRHVS	6872			
Db	1441	PVTHS	PL	LH	PRGFL	R	PSASL	PEEA	EASERSTEAP	PASPEGAGPPAAOCVPRHVS	1500			
QY	6873	LFYHQ	AGES	PEHG	AL	AF	GSRRH	PARR	HL	LKGGYIAGAL	PGLREPLMEHRVLESEAREE	6932		
Db	1501	LFYHQ	AGES	PEHG	AL	AF	GSRRH	PARR	HL	LKGGYIAGAL	PGLREPLMEHRVLESEAREE	1560		
QY	6933	QATLLA	KAP	FETAL	R	L	PAS	GTHLA	PGHSH	LSLHDSPT	PRPSSACEAQR	L	PSAPSGG	6992
Db	1561	QATLLA	KAP	FETAL	R	L	PAS	GTHLA	PGHSH	LSLHDSPT	PRPSSACEAQR	L	PSAPSGG	1620

Qy	6993	API	DMGH	POG	S	K	O	L	P	S	T	G	H	G	T	A	O	P	E	R	S	P	S	P	M	G	O	P	A	P	F	C	H	P	K	O	G	S	A	P	O	G	S	C	S	P	H	P	A		7052						
Db	1621	API	DMGH	POG	S	K	O	L	P	S	T	G	H	G	T	A	O	P	E	R	S	P	S	P	M	G	O	P	A	P	F	C	H	P	K	O	G	S	A	P	O	G	S	C	S	P	H	P	A		1680						
Qy	7053	VAP	CP	GS	F	P	PG	S	C	K	E	A	P	L	V	P	S	P	F	L	G	O	P	A	P	A	K	A	S	P	P	L	S	K	M	G	P	G	D	I	S	L	P	GR	K	P	G		7112								
Db	1681	VAP	CP	GS	F	P	PG	S	C	K	E	A	P	L	V	P	S	P	F	L	G	O	P	A	P	A	K	A	S	P	P	L	S	K	M	G	P	G	D	I	S	L	P	GR	K	P	G		1740								
Qy	7113	PCS	SP	GS	A	Q	A	S	S	Q	V	S	S	L	R	V	G	S	S	Q	V	T	E	P	G	P	S	L	D	A	E	G	M	T	O	E	A	E	D	I	S	D	T	P	T	L	Q	R	P	O	E	Q	A		7172		
Db	1741	PCS	SP	GS	A	Q	A	S	S	Q	V	S	S	L	R	V	G	S	S	Q	V	T	E	P	G	P	S	L	D	A	E	G	M	T	O	E	A	E	D	I	S	D	T	P	T	L	Q	R	P	O	E	Q	A		1800		
Qy	7173	TMR	X	F	S	I	G	R	G	G	Y	A	G	Y	T	F	A	G	D	A	G	M	L	C	Q	G	P	M	W	A	R	I	A	W	A	V	S	Q	S	E	E	E	E	E	E	E	E	A	R	A	S		7232				
Db	1801	TMR	X	F	S	I	G	R	G	G	Y	A	G	Y	T	F	A	G	D	A	G	M	L	C	Q	G	P	M	W	A	R	I	A	W	A	V	S	Q	S	E	E	E	E	E	E	E	E	A	R	A	S		1860				
Qy	7233	QSE	B	Q	Q	E	A	R	A	S	P	L	P	O	V	S	A	R	P	V	E	G	R	A	P	T	R	S	S	P	E	P	T	P	M	E	D	I	G	O	V	S	L	V	O	I	R	D	L	S	G	D	A	E	A		7292
Db	1861	QSE	B	Q	Q	E	A	R	A	S	P	L	P	O	V	S	A	R	P	V	E	G	R	A	P	T	R	S	S	P	E	P	T	P	M	E	D	I	G	O	V	S	L	V	O	I	R	D	L	S	G	D	A	E	A		1920
Qy	7293	DTI	S	L	D	I	S	E	V	D	P	A	Y	L	N	S	D	I	Y	K	I	P	F	F	M	I	P	R	K	V	P	K	S	A	Q	P	E	P	P	S	P	M	A	E	E	L	A	E	P	P	T		7352				
Db	1921	DTI	S	L	D	I	S	E	V	D	P	A	Y	L	N	S	D	I	Y	K	I	P	F	F	M	I	P	R	K	V	P	K	S	A	Q	P	E	P	P	S	P	M	A	E	E	L	A	E	P	P	T		1980				
Qy	7353	WPM	P	G	E	L	G	H	A	G	I	E	T	E	S	D	V	D	A	L	A	E	A	V	G	R	K	W	S	S	P	S	R	L	P	H	F	P	G	R	H	L	P	L	D	E	P	A	E	L		7412					
Db	1981	WPM	P	G	E	L	G	H	A	G	I	E	T	E	S	D	V	D	A	L	A	E	A	V	G	R	K	W	S	S	P	S	R	L	P	H	F	P	G	R	H	L	P	L	D	E	P	A	E	L		2040					
Qy	7413	GLR	E	R	V	K	A	S	V	E	H	I																																													

RESULT 6
US-10-093-463-72
; Sequence 72, Application US/10093463
; Publication No. US20030208039A1

GENERAL INFORMATION:
APPLICANT: Padigaru, Muralidhara
APPLICANT: Shenoy, Suresh
APPLICANT: Kekuda, Ramesh
APPLICANT: Gusev, Vladimir
APPLICANT: Pochart, Pascal
APPLICANT: Zhong, Mei
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter
APPLICANT: Smithson, Glennnda
APPLICANT: Guo, Xiaojia
APPLICANT: Gerlach, Valerie
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Ferenc
APPLICANT: Li, Li
APPLICANT: Zerhusen, Bryan
APPLICANT: Tchernev, Velizar
APPLICANT: Gangolli, Esha
APPLICANT: Vernet, Corine
APPLICANT: Pena, Carol
APPLICANT: Burgess, Catherine
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly
APPLICANT: Gorman, Linda
APPLICANT: Spaderna, Steven
APPLICANT: Voss, Edward
APPLICANT: Malyankar, Uriel
APPLICANT: Anderson, David
APPLICANT: Patturajan, Meera
APPLICANT: Miller, Charles
APPLICANT: Taupier, Raymond J. Jr.
TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypeptide
TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.
FILE REFERENCE: 21402-290A (Cura 590Ant)
CURRENT APPLICATION NUMBER: US/10/093,463
CURRENT FILING DATE: 2002-06-24
PRIOR APPLICATION NUMBER: 60/283,675
PRIOR FILING DATE: 2001-04-14
PRIOR APPLICATION NUMBER: 60/338,092
PRIOR FILING DATE: 2001-12-03
PRIOR APPLICATION NUMBER: 60/274,281
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/274,101
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/325,681
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/304,354
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/279,995
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 60/294,899
PRIOR FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 60/277,424
PRIOR FILING DATE: 2001-04-30
PRIOR APPLICATION NUMBER: 60/299,027
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: 60/309,198
PRIOR FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/281,194
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/274,194
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/274,849
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/330,380
PRIOR FILING DATE: 2001-10-19
PRIOR APPLICATION NUMBER: 60/275,235
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: 60/288,342
PRIOR FILING DATE: 2001-05-03
PRIOR APPLICATION NUMBER: 60/275,578
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 370

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 72
LENGTH: 4691
TYPE: PRT
ORGANISM: Homo sapiens
US-10-093-463-72

Query Match 25.5%; Score 10519.5; DB 15; Length 4691;
Best Local Similarity 51.1%; Pred. No. 0;
Matches 2367; Conservative 118; Mismatches 342; Indels 1807; Gaps 66;
QY 2595 MPLYNDSPHEISHKGRRTTLVLSIQRADAGIVRASSLVKSTASARLVKVPVFLKALD 2654
Db 1 MPLYNDSPHEISHKGRRTTLVLSIQRADAGIVRASSLVKSTASARLVKVPVFLKALD 60
QY 2655 DLSAEERGTALQCEVSDPEAHVVRKDGVLQPSDKYDFLHTACTRGLVHVDVSPEDAG 2714
Db 61 DLSAEERGTALQCEVSDPEAHVVRKDGVLQPSDKYDFLHTACTRGLVHVDVSPEDAG 120
QY 2715 LYTHVGSEETRARVRVHDLHVGITKRLKTMVELEGESCSPECVLSHESASDPAMTVGG 2774
Db 121 LYTHVGSEETRARVRVHDLHVGITKRLKTMVELEGESCSPECVLSHESASDPAMTVGG 180
QY 2775 KTVGSSSRFOATROGRKYILVVREAAPSDAGEVWFSVRLTSKSLIVRERPAAIKPLE 2834
Db 181 KTVGSSSRFOATROGRKYILVVREAAPSDAGEVWFSVRLTSKSLIVRERPAAIKPLE 240
QY 2835 DQWVAPGEDVELRCELSRAGTPVHMLKDKKAIKRSQKYDVVCEGTMAMLVIRGASLKDAG 2894
Db 241 DQWVAPGEDVELRCELSRAGTPVHMLKDKKAIKRSQKYDVVCEGTMAMLVIRGASLKDAG 300
QY 2895 EYTCVEASKSTASLHVBEKANCFTTELTNLQVEBKGTAVFTCKTEHPAAVTVWRKGLLE 2954
Db 301 EYTCVEASKSTASLHVBEKANCFTTELTNLQVEBKGTAVFTCKTEHPAAVTVWRKGLLE 360
QY 2955 LRASGHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSRQAQLVQGRVHHIIEDELDV 3014
Db 361 LRASGHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSRQAQLVQGRVHHIIEDELDV 420
QY 3015 DVQEGSSATFRCRISPANYEPVHWFDPKTPHLANELNEIDAOPGGYHVLTLRQALKDSG 3074
Db 421 DVQEGSSATFRCRISPANYEPVHWFDPKTPHLANELNEIDAOPGGYHVLTLRQALKDSG 480
QY 3075 TIYFAGDQASAAALRVTEKPSVFSRELTDATITEGEDTLVCETSTCDIPMCWTGDKGT 3134
Db 481 TIYFAGDQASAAALRVTEKPSVFSRELTDATITEGEDTLVCETSTCDIPMCWTGDKGT 540
QY 3135 LRGSARCOLSHEGHRQAQLITGATLQDSGRYKCEAGGACSSIVRVHARPVRFQALKDOL 3194
Db 541 LRGSARCOLSHEGHRQAQLITGATLQDSGRYKCEAGGACSSIVRVHARPVRFQALKDOL 600
QY 3195 EVLEGGAATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPQDSGRY 3254
Db 601 EVLEGGAATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPQDSGRY 660
QY 3255 SCSFGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAAPVWRKGSSETLRD 3314
Db 661 SCSFGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAAPVWRKGSSETLRD 720
QY 3315 GDRYCLRODGAMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPMPAHFIGRLRHQESI 3374
Db 721 GDRYCLRODGAMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPMPAHFIGRLRHQESI 780
QY 3375 EGATATLRCELSKAAPVWRKGRSLRQDGRHSRLRQDQAVCELOICGLAVADAGEYSCVC 3434
Db 781 EGATATLRCELSKAAPVWRKGRSLRQDGRHSRLRQDQAVCELOICGLAVADAGEYSCVC 840
QY 3435 GEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCWELSKVAPVWRKGPENLRDGRY 3494
Db 841 GEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCWELSKVAPVWRKGPENLRDGRY 900
QY 3495 ILRQEGTRCELOICGLAMADAGEYLCVCGEERTSATLIRALPARFIEDVKQNEAREGAT 3554
Db 3554 ILRQEGTRCELOICGLAMADAGEYLCVCGEERTSATLIRALPARFIEDVKQNEAREGAT 3554

Db 901 ILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTITRALPARFIEDVKNQAREGAT 960
Qy 3555 AVLQCELSAAPVEMWKGSETLRDGRYSLRQDGTKECELOIRGLAMADTGEYSCVCGQER 3614
Db 961 AVLQCELSAAPVEMWKGSETLRDGRYSLRQDGTKECELOIRGLAMADTGEYSCVCGQER 1020
Qy 3615 TSAMLTVRALPIKFTBGLRNEBATEGATAVLQCELSKAAPVEMWKGHETLRDGRHSLRQ 3674
Db 1021 TSAMLTVRALPIKFTBGLRNEBATEGATAVLQCELSKAAPVEMWKGHETLRDGRHSLRQ 1080
Qy 3675 DGARCELOIRGLVAEDDAGYLCWCGKERTSAMLTVRAMPSKEIEGLRNEEATEGATATLW 3734
Db 1081 DGARCELOIRGLVAEDDAGYLCWCGKERTSAMLTVRAMPSKEIEGLRNEEATEGATATLW 1140
Qy 3735 CELSKAAPVEMWKGHETLRDGRHSLRQDGRSCELOIRGLAVVDAGEYSCVCGQERTSAT 3794
Db 1141 CELSKAAPVEMWKGHETLRDGRHSLRQDGRSCELOIRGLAVVDAGEYSCVCGQERTSAT 1200
Qy 3795 LTVRALPARFIEDVKNQAREGATAVLQCELSKAAPVEMWKGSETLRDGRYSLRQDGT 3854
Db 1201 LTVRALPARFIEDVKNQAREGATAVLQCELSKAAPVEMWKGSETLRDGRYSLRQDGT 1260
Qy 3855 CELOIHGLSVADTGEYSCVCGQERTSATLTVR----- 3886
Db 1261 CELOIHGLSVADTGEYSCVCGQERTSATLTVRALPARFTQDLKTKBEGATATLOCELS 1320
Qy 3887 ----- 3886
Db 1321 KVAPVEMWKGSETLRDGRYSLRQDGTCELOIHDLVSADAGEYSCMCGQERTSATLTVR 1380
Qy 3887 ----- 3886
Db 1381 ALPARFTEGLRNEEAMEGATATLOCELSKAAPVEMWKGLEALRDGDKYSLRDGAVCELO 1440
Qy 3887 ----- 3886
Db 1441 IHGLAMADGVYSCVCGQERTSATLTVRALPARFIEDMRNOKATEGATVTLQCKLRKAAP 1500
Qy 3887 ----- 3886
Db 1501 VEMWKGPNLTKDGRYSLRQDGTSCLOIRGLVIADAGEYSCIQERTSATLTVRALPA 1560
Qy 3887 ----- 3886
Db 1561 RFIEDVRNHEATEGATAVLQCELSKAAPVEMWKGSETLRDGRYSLRQDGTCELOIRGL 1620
Qy 3887 ----- 3886
Db 1621 AVEDTGEYLCVCGQERTSATLTVRALPARFIDNMNTOAREGATATLHCELSKVAPVEMW 1680
Qy 3887 ----- 3886
Db 1681 KGPETLRDGRHSLRQDGRSCELOIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIE 1740
Qy 3887 ----- 3886
Db 1741 DVKNQAREGATAVLQCELSKAAPVEMWKGSETLRDGRYSLRQDGTCELOIHGLSVAD 1800
Qy 3887 ----- 3886
Db 1801 TGEYSCVCGQERTSATLTVRALPARFTQDLKTKBEGATATLOCELSKVAPVEMWKGPE 1860
Qy 3887 ----- 3886
Db 1861 TLRDGRYSLRQDGTCELOIHDLVSADAGEYSCMCGQERTSATLTVRDCHTLVMPHY 1920
Qy 3887 ----- 3886
Db 1921 FQLPGLKPEBTLYIQIPSPVILFTBGLRNEEAMEGATATLOCELSKAAPVEMWKGLE 1980
Qy 3887 ----- 3886
Db 1981 ALRDGDKYSLRDGAVCELOIHGLAMADGVYSSILPARFIEDMRNOKATEGATVTLQCKL 2040

Qy 3887 ----- 3886
Db 2041 RKAAPVEMWKGPNLTKDGRYSLRQDGTSCLOIRGLVIADAGEYSCICEQERTSATLT 2100
Qy 3887 ----- 3886
Db 2101 RALPARFIEDVRNHEATEGATAVLQCELSKAAPVEMWKGSETLRDGRYSLRQDGTCELO 2160
Qy 3887 ----- 3886
Db 2161 QIRGLAVEDTGEYLCVCGQERTSATLTITVRALPARFIDNMNTOAREGATATLHCELSKVA 2220
Qy 3887 ----- 3886
Db 2221 PVEMWKGPELTRDGRHSLRQDGTCELOIRGLSVADAGEYSCVCGQERTSATLTITRALP 2280
Qy 3887 ----- 3886
Db 2281 AKFTKGLRNEBATEGATAMLOCELSKVAPVEMWKGPELTRDGRYNLNRQDGTCELOIHG 2340
Qy 3887 ----- 3886
Db 2341 LSVADTGEYSCVCGQERTSATLTITVRALPARFIEDVKNQAREGATATLOCELSSEPTATV 2400
Qy 3921 WSKGGLQLOQANGREPRELQCGCTABLVLDQLOREDTGEYTCCTCGSQATSATLTITVTAAPVRF 3981
Db 2401 WSKGGLQLOQANGREPRELQCGCTABLVLDQLOREDTGEYTCCTCGSQATSATLTITVTAAPVRF 2460
Qy 3982 LRELQHOEVDGEGTAHLCCELSRAGASVEMWKGSLQLOPPCAKYQMVODGAAAEALLVRGVE 4041
Db 2461 LRELQHOEVDGEGTAHLCCELSRAGASVEMWKGSLQLOPPCAKYQMVODGAAAEALLVRGVE 2520
Qy 4042 QEDAGDYTCDTGHTQSNASLSVRPRPKFTRLOSLBQETGDIARLCCQLSDASGAVVQ 4101
Db 2521 QEDAGDYTCDTGHTQSNASLSVRGGR-----GAA-- 2549
Qy 4102 WLKGEVLEHAGPKYEMRSQGATRELLIHOLEAKDTGEYACVTCGQKTAASLRVTEPEVTI 4161
Db 2550 -----CGPQVRDAAQAGATRELLIHOLEAKDTGEYACVTCGQKTAASLRVTEPEVTI 2600
Qy 4162 VRGLVDAEVTDAEDVEFSCEVSRAGATGVQWCLQGLFLOSNEVTEVAVRDGRIHTLRKG 4221
Db 2601 VRGLVDAEVTDAEDVEFSCEVSRAGATGVQWCLQGLFLOSNEVTEVAVRDGRIHTLRKG 2660
Qy 4222 VTPEDAGTVSFHLGNHASSAQLTVRAPEVITLPLQDVOLSEGDQDASFQCRLSRASQGEA 4281
Db 2661 VTPEDAGTVSFHLGNHASSAQLTVRAPEVITLPLQDVOLR----- 2701
Qy 4282 RWALGGVPLQANEMNDITVEQGTLLHLLTLHKVTLEDAGTVSFHVGTCSSEALQKVT --- 4337
Db 2702 -----GVPLQANEMNDITVEQGTLLHLLTLHKVTLEDAGTVSFHVGTCSSEALQKVTBAVP 2756
Qy 4338 ----- 4337
Db 2757 CLVRGLQNVDFAGEVATFSCEDGPQSAIAVRDGIHFHSLMLSGLGADSGTVIFRAGPLV 2816
Qy 4338 ----- 4337
Db 2817 STAKLLIKDPVVEVVSAMQDLAVEEGSABELLCOYSRPVQATWKMDBREVTDHGRVIE 2876
Qy 4338 -----AKNTVVRGLENVEALEGEALFE 4360
Db 2877 QDMNVARLTFRPALPCDSGIYSCEAAGTRVALLQVQAKNTVVRGLENVEALEGEALFE 2936
Qy 4361 COLSQOPEVAATHLLDDEPVRTSENABEVFPENGLRHLLLLKNLRPODSKRVTFLAGDMV 4420
Db 2937 COLSQOPEVAATHLLDDEPVRTSENABEVFPENGLRHLLLLKNLRPODSKRVTFLAGDMV 2996
Qy 4421 TSAFLTVR-----GWRLEILEPLKNAAVRAGAQAARFTCTLSEAVPVGEASWYINGAA 4472
Db 2997 TSAFLTVRGCAVLVQWRLEILEPLKNAAVRAGAQAARFTCTLSEAVPVGEASWYINGAA 3056

QY 4473 VOPDDSDMTVTADGSHQALLLSAOPHHAGEVTFACRDVAVASARLTVLGLPDPPEAEVV 4532
DB 3057 VOPDDSDMTVTADGSHHALLLSAOPHHAGEVTFACRDVAVASARLTVLGLPDPPEAEVV 3116
QY 4533 AHSSHTVTLSSWAAPNSDGGGLCGYRVVEKEGATQWRLCHELVPGPECVVDGLAPGETY 4592
DB 3117 ARSSHTVTLSSWAAPNSDGGGLCGYRVVEKEGATQWRLCHELVPGPECVVDGLAPGETY 3176
QY 4593 RFRVAAGVPGVAGEPVHLPTQVRLAEPKPVPPQPSAPESROVAAGEDVLSLEVVABAG 4652
DB 3177 RFRVAAGVPGVAGEPVHLPTQVRLAEPKPVPPQPSAPESROVAAGEDVLSLEVVABAG 3235
QY 4653 EVIWHKMERIOPGRFEVVSQGRQOVLVIKGTAEQDQYHCGLAQGSICPAATFOVA 4712
DB 3236 EVIWHKMERIOPGRFEVVSQGRQOVLVIKGTAEQDQYHCGLAQGSICPAATFOVA 3295
QY 4713 LSPASVDEAPQSLPPEAAQEGDHLLEALARKRMSREPTLDSISELPEEDGRSQRPL 4772
DB 3296 LSPASVDEAPQSLPPEAAQEGDHLLEALARKRMSREPTLDSISELPEEDGRSQRPL 3355
QY 4773 QBAEVAFDLSRGYSTADELARTGADLSHTSSDDESAGTSPSLVLYLKAGRPQTSPLA 4832
DB 3356 QBAEVAFDLSRGYSTADELARTGADLSHTSSDDESAGTSPSLVLYLKAGRPQTSPLA 3415
QY 4833 SKVGAP-----AAPSVPQOQOQPL 4852
DB 3416 SKVSPNLAKEFRPTPRAGRSLLGFVGADPAFPGSERSARCTRCAAPPRESUKREP- 3474
QY 4853 AAVRPPLGDLSTKDLG-----DPSMD-----KAAPK 4884
DB 3475 ASCLP--GAMEAVELARKLQEBATCSICLDYFTDPVMTTCGHNFCEACIQLSWEXARGK 3532
QY 4885 GYKVRK-----EMKQEGPMPSHTFGDTEAOGDARLECVVASKADVARWLKDG 4935
DB 3533 GRKKEGSPPCDECREMSQPNLLN-----RLLTQVAEMA----- 3568
QY 4936 VELTDGRHHIDQLGDGTCSLIAGLDRADAGCYTCOVSNK-----FGQVTHSACVVVSGS 4991
DB 3569 -----QOH-----PGLQKOD-----LCQEHHEPLKLFCKQKQSPICVVCRE 3604
QY 4992 ESEAESGGELDDAFRAARLHLRFTKSPAESVDBELFLSADEG----- 5038
DB 3605 SRE-----HRLHVL-----PABEAVQYKLLBEDMEYLREQITRGN 3643
QY 5039 --PAPPEPADWQ--TYRDEHPCITRPERAL-----TEARQAVTRPQEMF 5079
DB 3644 LQAREEQSLAEWQGVKERRERIVLEFEKMNLYLVEEQRLIQALETBEEETASRLRESV 3703
QY 5080 ATL---GIGVEIKLVE-----QGPFR-----VEMCISK-- 5104
DB 3704 ACLDRQGHSLLELLLIQLBERSTQGPLQMLQDMKEPLSRAALLVLIHGMNLVEFFPVVSLP 3763
QY 5105 -----ETPAPVVPPEPLPSLLTSDAAP-----VFLTE 5131
DB 3764 SPLYLIAKHAQTOLGPGTTPDEPCTPLPISPP-PRFS--TEDVVPDATSAYPYLLIYE 3820
QY 5132 LQNOEVQDGY-----VSFDCVVTGQPMPSVRFKDGKLLBEDDHYMINEDOQG 5180
DB 3821 SRQRRYLSSPBGSGFCSDKDFVAVPCAV-GQTA-----FSSCR-----HYWEVGMNIT 3868
QY 5181 GHQLIITAVPADMG-----VYRCLAESNGVSGTAEKLRVDLTSTDYDTAADATSSSYF 5236
DB 3869 GDALWALGVCRNDRKDRVPKC--PENGFW-----VOLSK-----GTKYLSTF 3911
QY 5237 SAQGYLSRE---QBSTSTTDEGQLPVVVEBLRLQVAPGTRLAKFOLKVKVGPAPRLY 5293
DB 3912 SALTPVLMLEPPSHNGIFLDPEAG-----EVSFYSVSDGSHLHTYSQAT--FGPQLP 3962
QY 5294 WFKQOQITASAHIRMTGKILHTLEIISVTRDSGQVAAVYTSNANGAAYSSARLLVRGP 5353
DB 3963 PFCLGAP-----KSGQWVSTVTMAGV-KDLATRTGAVVTPALGA----- 4001
QY 5354 DEPEKPADSVHQLVPPRMRLERFTPKKVKGSSITFFSVKVEGRPVPTVHMLREBERGV 5413

DB 4002 -----YAPSATETQS-----PAP-----WSFRAPE--- 4021
QY 5414 LWIGPTFTCYTVASSAQHSLVLLDVGRHQHGYTTCIASNAAGQALCSASLHVSGLPKVE 5473
DB 4022 ---PEHFG--VPSLAPRSA-----RACAAAPGVPGSPRAA 4051
QY 5474 BOEKYKEALISTFLOGTTOAISQAQLETSAPADLGGQORKEEPLAAKEALHLSLAVGTE 5533
DB 4052 EAARRRPADSTAFPL-SVRAMAA-----PDLSTNLQEEATCA----- 4087
QY 5534 EFLQKLTSQITEMVSAKITQAKLOVPGGSDSDSTPSAS--PRGRSRPSSSI----- 5585
DB 4088 ICLDYFTDPVMTDCGHNFCEIRCWQOPEARTRAPSAACPRGTCGPTARLLRWPRW 4147
QY 5586 QSSSESDGDGARGEIFDIYVVVTDYPLGAE-QDAITL-----REGQYVEVLDAAHPL 5638
DB 4148 RGACTRRRRRRR-----VPAHREPLAAFCGDELRLCAACERSGEH----- 4189
QY 5639 RWL-----VRYKPTKSSPSRQ-----WSPAYLDRRLK-LSPE----- 5671
DB 4190 -WAHRVGRCTRPTKTSRPLEAGTMAANETLLSGAKLEKSLHLRKQMQDALLFOAQDET 4248
QY 5672 ---WCAEAPFPFGAEVSEDEYKARLSSVIOELLSSQAFVEELOFLQSHHL---OHL 5724
DB 4249 CVLWQABDGGGQRQNVLRFE---RLRLLAECGTTAAAEAGEEELKQSAHLAELE 4305
QY 5725 RCPHVPIAVAGOKA 5738
DB 4306 R--PLPAAACAGAAA 4317

RESULT 7

US-10-093-463-74
; Sequence 74, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zethusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esba
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypeptide
; FILE OF INVENTION: Encoding the Antigens, and Methods of Use.
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675


```
/ PRIOR FILING DATE: 2001-04-14
/ PRIOR APPLICATION NUMBER: 60/338,092
/ PRIOR FILING DATE: 2001-12-03
/ PRIOR APPLICATION NUMBER: 60/274,281
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/274,101
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/325,681
/ PRIOR FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: 60/304,354
/ PRIOR FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: 60/279,995
/ PRIOR FILING DATE: 2001-03-30
/ PRIOR APPLICATION NUMBER: 60/294,899
/ PRIOR FILING DATE: 2001-05-31
/ PRIOR APPLICATION NUMBER: 60/287,424
/ PRIOR FILING DATE: 2001-04-30
/ PRIOR APPLICATION NUMBER: 60/299,027
/ PRIOR FILING DATE: 2001-06-18
/ PRIOR APPLICATION NUMBER: 60/309,198
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/281,194
/ PRIOR FILING DATE: 2001-04-04
/ PRIOR APPLICATION NUMBER: 60/274,194
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/274,849
/ PRIOR FILING DATE: 2001-03-09
/ PRIOR APPLICATION NUMBER: 60/330,380
/ PRIOR FILING DATE: 2001-10-18
/ PRIOR APPLICATION NUMBER: 60/275,235
/ PRIOR FILING DATE: 2001-03-12
/ PRIOR APPLICATION NUMBER: 60/288,342
/ PRIOR FILING DATE: 2001-05-03
/ PRIOR APPLICATION NUMBER: 60/275,578
/ PRIOR FILING DATE: 2001-03-13
/ NUMBER OF SEQ ID NOS: 370
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 74
/ LENGTH: 4675
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-093-463-74

Query Match      25.5%; Score 10506; DB 15; Length 4675;
Best Local Similarity 56.1%; Pred. No. 0;
Matches 2252; Conservative 45; Mismatches 129; Indels 1586; Gaps 35;

Qy      2595 MPLYNDSFHEISHKGRHTLVLSKIQRADAGIVRASSLKVSTSALEVRVPVFLKALD 2654
Db      1 MPLYNDSFHEISHKGRHTLVLSKIQRADAGIVRASSLKVSTSALEVRVPVFLKALD 60

Qy      2655 DLSAERTGLALQCRVSPPEAHVVRKDGVLGPSDKYDFLHTAGTGLVVDVSPEDAG 2714
Db      61 DLSAERTGLALQCEVSPPEAHVVRKDGVLGPSDKYDFLHTAGTGLVVDVSPEDAG 120

Qy      2715 LYTHVGSSETRARVRVHDLHVGIKRLKTMVLEGESCSPECVLSHESADPAMWTVGG 2774
Db      121 LYTHVGSSETRARVRVHDLHVGIKRLKTMVLEGESCSPECVLSHESADPAMWTVGG 180

Qy      2775 KTVGSSSFQATROGRKYLIVVREAPSDAGEVVFVSRGLTSKASLIIVRPPAIIKPLE 2834
Db      181 KTVGSSSFQATROGRKYLIVVREAPSDAGEVVFVSRGLTSKASLIIVRPPAIIKPLE 240

Qy      2835 DQWAPGSDVELRCELSRAGTPVHVKDKAIRKSQKDYDVVCEGTMAMLVIRGASLKDAG 2894
Db      241 DQWAPGEDVELRCELSRAGTPVHVKDKAIRKSQKDYDVVCEGTMAMLVIRGASLKDAG 300

Qy      2895 EYTCVEASKSTASHVEKANCFTTELTNLQVEEKGTAFTCKTEHPAATVTVWRKGLLE 2954
Db      301 EYTCVEASKSTASHVEKANCFTTELTNLQVEEKGTAFTCKTEHPAATVTVWRKGLLE 360

Qy      2955 LRASGKHQPSQEGTLRLTITISALEKADSDTYTCDIQGASRAQLLVQRRVHIIEBDEV 3014
```


QY 3887 ----- 3886
Db 1501 VEWKGPNTLKDGDRYSLKQDGTSCELQIRGLVIADAGEYSCICEQERTSATLTVRALPA 1560
QY 3887 ----- 3886
Db 1561 RFIEDVRNHEATEGATAVLOCELSKAAPVWWRKGSSETLRDGRYSLRQDGTCELRQIRGL 1620
QY 3887 ----- 3886
Db 1621 AVEDTGEYLCVCGQERTSATLTVRALPARFIDNMTNQAREGATATLHCELSKVPVWR 1680
QY 3887 ----- 3886
Db 1681 KGPETLRDGRHSRLRQDSRCELRQIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIE 1740
QY 3887 ----- 3886
Db 1741 DVKNQAREGATAVLOCELSKAAPVWWRKGSSETLRDGRYSLRQDGTCELRQIRGLSVAD 1800
QY 3887 ----- 3886
Db 1801 TGEYSCVCGQERTSATLTVRALPARFTQDLTKKEASEGATATLQCELSKVPVWWRKGSPE 1860
QY 3887 ----- 3886
Db 1861 TLRDGRYSRLKQDGTCELRQIRGLHLSVADAGEYSCVCGQERTSATLTVRDCHTLHVMPHY 1920
QY 3887 ----- 3886
Db 1921 POLPGLKPEBETLIYI QIPSPVILFTEGLRNEAMEGATATLQCELSKAAPVWWRKGLE 1980
QY 3887 ----- 3886
Db 1981 ALRDGDKYSLRQDGAVALCELRQIRGLAMADNGVYSSLPARFIEDMRNQKATEGATVTLQCKL 2040
QY 3887 ----- 3886
Db 2041 RKAAPVWWRKGPNTLKDGDRYSRLKQDGTSCELQIRGLVIADAGEYSCICEQERTSATLT 2100
QY 3887 ----- 3886
Db 2101 RALPARFIEDVRNHEATEGATAVLOCELSKAAPVWWRKGSSETLRDGRYSRLRQDGTCELR 2160
QY 3887 ----- 3886
Db 2161 QIRGLAVEDTGEYLCVCGQERTSATLTVRALPARFIDNMTNQAREGATATLHCELSKVA 2220
QY 3887 ----- 3886
Db 2221 PVWWRKGPETLRDGRHSRLRQDGTCELRQIRGLSVADAGEYSCVCGQERTSATLTIRALP 2280
QY 3887 ----- 3886
Db 2281 AKFTKGLRNEATEGATAMLOCELSKVPVWWRKGPETLRDGRYSLRQDGTCELRQIRGLH 2340
QY 3887 ----- 3886
Db 2341 LSVADTGEYSCVCGQERTSATLTVRALPARFIDNMTNQAREGATATLHCELSKVA 2400
QY 3922 WSKGLOLQANGRRPRLOGCTAEVLVQDLQREDTGEYTCGSOATSATLTVRALPARF 3981
Db 2401 WSKGLOLQANGRRPRLOGCTAEVLVQDLQREDTGEYTCGSOATSATLTVRALPARF 2460
QY 3982 LRELQHEVDEGGTAHLCELSRAGASVWWRKGSQLPFCAPVQVODGAAELVVRGVE 4041
Db 2461 LRELQHEVDEGGTAHLCELSRAGASVWWRKGSQLPFCAPVQVODGAAELVVRGVE 2520
QY 4042 QEDAGDYTCDTGHTQSMASLSVRPRPKFTRQLQSLQEQETGDIARLCCQLSDAESGAVVQ 4101
Db 2521 QEDAGDYTCDTGHTQSMASLSVRGR -----GAA-- 2549

QY 4102 WLKGEVHLAGPKYEMRSQCATRELLIHOLEAKDTGEYACVTGGQKTAASLVRTEBEVTI 4161
Db 2550 -----CGPQVRDAAQCATRELLIHOLEAKDTGEYACVTGGQKTAASLVRTEBEVTI 2600
QY 4162 VRGLVDAEVTADVEDFEFSCEVSRAGATGVQWCLQGLPLOSNEVTEVAVRDGRHITLRLKG 4221
Db 2601 VRGLVDAEVTADVEDFEFSCEVSRAGATGVQWCLQGLPLOSNEVTEVAVRDGRHITLRLKG 2660
QY 4222 VTPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPIQDVOLSEGQDASFOCRLSRASGOBA 4281
Db 2661 VTPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPIQDVQLR ----- 2701
QY 4282 RHALCGVPLOANEMNDITVEQGTLLHLLTKHTLEDAGTVSFHVGTCSSQAOLKVT- 4337
Db 2702 -----GVPLQANEMNDITVEQGTLLHLLTKHTLEDAGTVSFHVGTCSSQAOLKVTAVP 2756
QY 4338 ----- 4337
Db 2757 CLVRLQNVDPAGEVATFSCEDGPQSAIAVRDGI FHSIMLSGLGVADSGTVI PRAGPLV 2816
QY 4338 ----- 4337
Db 2817 STAKLLIKDPVVEVVSAMQDLAVERGSAELLCOYSRPVQATWKMDEREVHTDGHVITE 2876
QY 4338 -----AKNTVVRLGLENVEALEGGEALFE 4360
Db 2877 QDMNVARLTFRPPALPCDSGIYSCAAAGTRVVALQVQAKNTVVRLGLENVEALEGGEALFE 2936
QY 4361 COLSPEVAANTWLLDDPVRTSENAEVVFFENGRLHLLLLKNLRPODSCRVTFLAGDMV 4420
Db 2937 COLSPEVAANTWLLDDPVRTSENAEVVFFENGRLHLLLLKNLRPODSCRVTFLAGDMV 2996
QY 4421 TSAFILTVR-----GWRLEILEPLKNAAVRAGAQAARFTCTLSBAVPVGEASWYINGAA 4472
Db 2997 TSAFILTVRDCAVLVQGWRLLEILEPLKNAAVRAGAQAARFTCTLSBAVPVGEASWYINGAA 3056
QY 4473 VOPDDSDMTVADGSHOALLRSQAOPHHAGEVTFACRDVAASARLTVLGLPPEDEAVV 4532
Db 3057 VOPDDSDMTVADGSHHALLRSQAOPHHAGEVTFACRDVAASARLTVLGLPPEDEAVV 3116
QY 4533 AHSHTVTLSWAA PMSDGGGLCGYRVEKCATQWRLCHELVGPECVVDGLAPGETY 4592
Db 3117 AHSHTVTLSWAA PMSDGGGLCGYRVEKCATQWRLCHELVGPECVVDGLAPGETY 3176
QY 4593 RFRVAAGVPVGAPEVHLPTVRLAEPKPPVPPQPSAPESRQVAAGEDVSLLEVVAEAG 4652
Db 3177 RFRVAAGVPVGAPEVHLPTVRL-EPKPPVPPQPSAPESRQVAAGEDVCLLEVVAEAG 3235
QY 4653 EVIHHKGMERTOPGGRFEVVSQGRQOOLVIKGFTHAEDQGEYHCGLAQSGICPAAATFQVA 4712
Db 3236 EVIHHKGMERTOPGGRFEVVSQGRQOOLVIKGFTHAEDQGEYHCGLAQSGICPAAATFQVA 3295
QY 4713 LSPASVDEAPPSLPPEAAQEGDLHLWEALARKRMSREPTLDSI SELPEBDDGRSRLP 4772
Db 3296 LSPASVDEAPPSLPPEAAQEGDLHLWEALARKRMSREPTLDSI SELPEBDDGRSRLP 3355
QY 4773 QEAEEVAPDLSEGYSTADELARTGADLSHTSSDDDESRAAGTSPSLVTLKACRPGTSPUA 4832
Db 3356 QEAEEVAPDLSEGYSTADELARTGADLSHTSSDDDESRAAGTSPSLVTLKACRPGTSPUA 3415
QY 4833 SKVGAP-----AAPSVPKQOOQEP 4852
Db 3416 SKVSPNIAKERRFPTPRAGRSLGLFVGADPAFPGSGRSARCTRRCAAPPRESLKRFP- 3474
QY 4853 AAVRPPLGLSTKDLG-----DPSMD-----KAAVKIQ-----AAPK 4884
Db 3475 ASCLP--GAMEAVELARKLQEBATCSICLDVFTDPVMTTCGHNFRCRACIQLSWEKARGKK 3532
QY 4885 GYKVRK-----EMKQKQGPMSFHTFGDTEAQVGDALRLCEYVASKADVARMLKDG 4935
Db 3533 GRRKEKSGFPCECREMSPQNLNPN-----RLLTKVAEMA----- 3568
QY 4936 VELTDGRHHHIDQLGDGTCSLLIAGLRADAGCYTCQVSNK-----FGQVTHSACVVVSGS 4991


```
QY 2942 PAATVTRKGLLELRASCKHOPQSGELTLRLTISALEKADSTYTCIDIGQASRAQLLVQ 3001
Db 901 PAATVTRKGLLELRASCKHOPQSGELTLRLTISALEKADSTYTCIDIGQASRAQLLVQ 960
QY 3002 GRRVHIIEDELDVDVQEGSSATFCRISIPANYPVHWFELDKTPLHANELNEIDAOPGGVH 3061
Db 961 GRRVHIIEDELDVDVQEGSSATFCRISIPANYPVHWFELDKTPLHANELNEIDAOPGGVH 1020
QY 3062 VLTLRQALKDSTGIYFAGDQASAAALRVTEKPSVFSRELTDAITTEGEDTLVCEST 3121
Db 1021 VLTLRQALKDSTGIYFAGDQASAAALRVTEKPSVFSRELTDAITTEGEDTLVCEST 1080
QY 3122 CDIPWCWTQDKTILGARSARCQLSHEGHRALQITATLQDSGRYKCEAGGACSSIVRVH 3181
Db 1081 CDIPWCWTQDKTILGARSARCQLSHEGHRALQITATLQDSGRYKCEAGGACSSIVRVH 1140
QY 3182 ARPVRFOALKDLEVLGGATLRCVLSVAAPVKWCYGNVLRPGDKYSLRQEGAMLEL 3241
Db 1141 ARPVRFOALKDLEVLGGATLRCVLSVAAPVKWCYGNVLRPGDKYSLRQEGAMLEL 1200
QY 3242 VVRNLRPDQSGRYSCSFGDQTTSATLTVTALPAQFIGLKRKEATEGATATLRCELSKTA 3301
Db 1201 VVRNLRPDQSGRYSCSFGDQTTSATLTVTALPAQFIGLKRKEATEGATATLRCELSKAA 1260
QY 3302 PVWRKGSSETLRDGRYCLRDGAMCELOIRGLAMVDAABEYSCVCGERTSASLTIRPMP 3361
Db 1261 PVWRKGSSETLRDGRYCLRDGAMCELOIRGLAMVDAABEYSCVCGERTSASLTIRPMP 1320
QY 3362 AHFTGLRHQSIIEGATATLRCELSKAAPVWRKGRSLRGDGRHSLRDGAVCELOICG 3421
Db 1321 AHFTGLRHQSIIEGATATLRCELSKAAPVWRKGRSLRGDGRHSLRDGAVCELOICG 1380
QY 3422 LAVADAGEYSCVCGERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCELSKVAPVEW 3481
Db 1381 LAVADAGEYSCVCGERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCELSKVAPVEW 1440
QY 3482 RKGPENLRDGRYILRQSGTCELOICGLAMADAGEYLCVCGERTSATLTIRALPARFI 3541
Db 1441 RKGPENLRDGRYILRQSGTCELOICGLAMADAGEYLCVCGERTSATLTIRALPARFI 1500
QY 3542 EDVKNQAREGATAVLQCELSAAAPVWRKGSSETLRDGRYSLRDGDKCELOIRGLAWA 3601
Db 1501 EDVKNQAREGATAVLQCELSAAAPVWRKGSSETLRDGRYSLRDGDKCELOIRGLAWA 1560
QY 3602 DTGEYSCVCGERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSSKAPVWVKGH 3661
Db 1561 DTGEYSCVCGERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSSKAPVWVKGH 1620
QY 3662 ETLRDGRHSLRDGARGCELOIRGLVADAGEYLCWCKERTSAMLTVRAMPSPKIEGLR 3721
Db 1621 ETLRDGRHSLRDGARGCELOIRGLVADAGEYLCWCKERTSAMLTVRAMPSPKIEGLR 1680
QY 3722 NEEATEGDTATLWCELSKAAAPVWRKGHETLRDGRHSLRDGSGRCELOIRGLAVDAGE 3781
Db 1681 NEEATEGDTATLWCELSKAAAPVWRKGHETLRDGRHSLRDGSGRCELOIRGLAVDAGE 1740
QY 3782 YSCVCGERTSATLTVRALPARFIEDVKNQAREGATAVLQCELSKAAAPVWRKGSSETLR 3841
Db 1741 YSCVCGERTSATLTVRALPARFIEDVKNQAREGATAVLQCELSKAAAPVWRKGSSETLR 1800
QY 3842 CGDRYSLRDGTRCELOIRGLSVADTGYSCVCGERTSATLTVRAPQVFPREPLQSLQA 3901
Db 1801 CGDRYSLRDGTRCELOIRGLSVADTGYSCVCGERTSATLTVRALPARFTQDLKTKEA 1860
QY 3902 EEGSTATLQCELSSEPTATVWWSKGLQLOANGRPREPLOGCTAEILVLODLOREDTGYTC 3961
Db 1861 SEGATATLQCELSK--VAPVWRKGPETLRDGRVSLKQDGTGRCELOIRGLSVADAGEYSC 1919
QY 3962 TCGSOATSATLTVAAPVRLRELQHQEBVDEGTAHLCCELSRAGASVWRKGSGLQFPFC 4021
Db 1920 MCGQERTSAMLTVRALPARFTEGLRNEEAMEGATATLQCELSKA--APVWRKGLEALRDG 1978
```

```
QY 4022 AKYQWQDGAABELLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRQLQSEQT 4081
Db 1979 DKYSLRQDGAAYCELCQIHGLAMADNGVYSCVCGERTSATLTVRALPARFIEDMRNOKATE 2038
QY 4082 GDIALRCOLSDAESSGAVVOMLKEGVELHAGPKYEMRSOGATRELLIHOLEBAKDTGEVAC 4141
Db 2039 GATVTLQCKLRKA--APVWRKGPNTLKGDRYSLKQDGTSCELQIRGLVADAGEYSC 2095
QY 4142 VTGQKTAASLRVTEPEVTIIVRGLVDAFVTADEDEVEFSCEVSRAGATGVQMCLQPLQOS 4201
Db 2096 ICEQERTSATLTVRALPARFIEDVRNHEATEGATAVLQCELSKAAAP--VWRKGSSETLRD 2153
QY 4202 NEVTEVAVRDCRIHTLRLKGVTPEDAGTVSHLGNHASSAQLTVRAPEVTILEPQDVQL 4261
Db 2154 GDYSLR--QDGTRELCQIRGLAVEDTGEYLCVCGERTSATLTVRALPARFIDNMNQB 2212
QY 4262 SEGQDASFOCRLSRASQOEAWALGGVPLQANEMNDITVEQGTLLHLLTLHKVTLLEDAGTV 4321
Db 2213 REGATATLHCELSKAAAPVE--WRKGRSLRGDRHSLR--QDGAAYCELOICGLAVADAGEY 2269
QY 4322 SFHVGTCSESAOLKV 4336
Db 2270 SCVCGERTSATLTV 2284

RESULT 9
US-09-858-664A-2
; Sequence 2, Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, KINASE PROTEINS, AND USES
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-2

Query Match 20.4%; Score 8423; DB 9; Length 1665;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6356 PPSQVITIEDVQAQTGGTAQPEAIIIEGDPQPSVTWYKDSVQLVDSSTRLSQQOEGTYSILV 6415
Db 53 PPSQVITIEDVQAQTGGTAQPEAIIIEGDPQPSVTWYKDSVQLVDSSTRLSQQOEGTYSILV 112
QY 6416 LRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGDNPEPDSKQSHRRKHLHSPFYEVKEBIG 6475
Db 113 LRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGDNPEPDSKQSHRRKHLHSPFYEVKEBIG 172
QY 6476 RGVEGFVKRVQHKGNKILCAAKFIPLASRTBAQAYRERDILAAALSHPLVTLGLDQFETRK 6535
Db 173 RGVEGFVKRVQHKGNKILCAAKFIPLASRTBAQAYRERDILAAALSHPLVTLGLDQFETRK 232
QY 6536 TLILILELCSSEELLDRLRYKGVVTEAEVKYIIOQLVEGLYHLHSHGVHLHIDIKPSNILM 6595
Db 233 TLILILELCSSEELLDRLRYKGVVTEAEVKYIIOQLVEGLYHLHSHGVHLHIDIKPSNILM 292
QY 6596 VHPAREDIKICDFGAQNTIPAELOFQSGYSPFVSPFPIIQNVPVSEASDIWAMGVISYL 6655
Db 293 VHPAREDIKICDFGAQNTIPAELOFQSGYSPFVSPFPIIQNVPVSEASDIWAMGVISYL 352
QY 6656 SLTCSPPAGESDRTALLNLEGRVSWSSPMAHLSEDAKDFIKATLQRAQPARSAQC 6715
```



```
QY 6836 AEASERSTEAPAPPASPEGAGPPAAQCGVPRHSVIRSLFYHQAGSPBEHGALAPGSRHRP 6895
DB 533 AEASERSTEAPAPPASPEGAGPPAAQCGVPRHSVIRSLFYHQAGSPBEHGALAPGSRHRP 592
QY 6896 ARRRHLLKGGYTAGALPGLRPLMBHRVLEBEAAREEQATLLAKAPSPETALRLPASGTH 6955
DB 593 ARRRHLLKGGYTAGALPGLRPLMBHRVLEBEAAREEQATLLAKAPSPETALRLPASGTH 652
QY 6956 LAPGSHSLEHDSPTSPSSACGEAQLPSAPSGGAPIRDMGHPQGSQKLPSTGGHPG 7015
DB 653 LAPGSHSLEHDSPTSPSSACGEAQLPSAPSGGAPIRDMGHPQGSQKLPSTGGHPG 712
QY 7016 TAQPERPSPDSWGPAPFCHPKQSAQOEGCSPPHAPVAPCPGSPFPSCKEAPLVPSS 7075
DB 713 TAQPERPSPDSWGPAPFCHPKQSAQOEGCSPPHAPVAPCPGSPFPSCKEAPLVPSS 772
QY 7076 PFLGQOQAPPAPAKASPLPDSKMGPGDISLQCRPKPGPCSSPGSASQASSQVSSLRVGS 7135
DB 773 PFLGQOQAPPAPAKASPLPDSKMGPGDISLQCRPKPGPCSSPGSASQASSQVSSLRVGS 832
QY 7136 SQVTEPGSLDAEGWTQEAEDLSPTTLQRPQEQATMRKPSLGGRGYAGVAGTGTA 7195
DB 833 SQVTEPGSLDAEGWTQEAEDLSPTTLQRPQEQATMRKPSLGGRGYAGVAGTGTA 892
QY 7196 FCGDAGGMLQGPMPWARIATAWVSQEEBQBARAESQEEQBARAESPLPQVSGARPVP 7255
DB 893 FCGDAGGMLQGPMPWARIATAWVSQEEBQBARAESQEEQBARAESPLPQVSGARPVP 952
QY 7256 EVGRAPTSSPPTPWEIDIGVSLQVIRDLSDGAADATISLIDISEVDPAAYLNLSLDYDI 7315
DB 953 EVGRAPTSSPPTPWEIDIGVSLQVIRDLSDGAADATISLIDISEVDPAAYLNLSLDYDI 1012
QY 7316 KYLPPEFMIFRKVPKSAQPEPPSPMAEELAEFPPTWPGELGPHAGLEITESESDVD 7375
DB 1013 KYLPPEFMIFRKVPKSAQPEPPSPMAEELAEFPPTWPGELGPHAGLEITESESDVD 1072
QY 7376 ALLAAEAAGVRKWKNSPRLSPLHFGRLPLDEPAELGLRVRKVASVEHISRLKGRPEG 7435
DB 1073 ALLAAEAAGVRKWKNSPRLSPLHFGRLPLDEPAELGLRVRKVASVEHISRLKGRPEG 1132
QY 7436 LEKEGPPRRKPKGLASFRISGLKSWDRAPTFRLSDETVLQGSVTLACQVSAQAQAAT 7495
DB 1133 LEKEGPPRRKPKGLASFRISGLKSWDRAPTFRLSDETVLQGSVTLACQVSAQAQAAT 1192
QY 7496 WSKOGAPLESSSRVLISATLKNFOLLTILVVAEDLGVYTCVSNALGTVTTTGVLRKAE 7555
DB 1193 WSKOGAPLESSSRVLISATLKNFOLLTILVVAEDLGVYTCVSNALGTVTTTGVLRKAE 1252
QY 7556 RPSSSPCPDIGHVYADGVLVWKPVESYGPVTYIVQCSLEGGSWTTLASDIPDCCYLTSK 7615
DB 1253 RPSSSPCPDIGHVYADGVLVWKPVESYGPVTYIVQCSLEGGSWTTLASDIPDCCYLTSK 1312
QY 7616 LSRGTYTFRACVSKAGMGYPSSPEOVLLGGPSHLASEEBSQGRSAQPLPSTKTFAPQ 7675
DB 1313 LSRGTYTFRACVSKAGMGYPSSPEOVLLGGPSHLASEEBSQGRSAQPLPSTKTFAPQ 1372
QY 7676 TQIQGRFSVVRQCEKASGRALAAKIIPIYHPKDKTAVLREYALKGLRHPHQAQHAAY 7735
DB 1373 TQIQGRFSVVRQCEKASGRALAAKIIPIYHPKDKTAVLREYALKGLRHPHQAQHAAY 1432
QY 7736 LSPRLHLVILELCSGPPELLPCLAEASVSESVKQYLQWMLSATQYLNQHTLHLDLRSE 7795
DB 1433 LSPRLHLVILELCSGPPELLPCLAEASVSESVKQYLQWMLSATQYLNQHTLHLDLRSE 1492
QY 7796 NMIITEYNLLKVVDLGNQSLQSEKVLPSDKFQDYLETWAPPELLGGQAVPOTDIIWAIGV 7855
DB 1493 NMIITEYNLLKVVDLGNQSLQSEKVLPSDKFQDYLETWAPPELLGGQAVPOTDIIWAIGV 1552
QY 7856 TAFIMLSAEYPVSSSEGARDLQRLKGLVRLSRCYAGLSGGGAVAFRLSTLCAQPWGRPCA 7915
DB 1553 TAFIMLSAEYPVSSSEGARDLQRLKGLVRLSRCYAGLSGGGAVAFRLSTLCAQPWGRPCA 1612
```

```
QY 7916 SSCLOCPMLTBEGPACSRPAPVTFPTARLRFVVRNREKRALLYKRNLAQVR 7968
DB 1613 SSCLOCPMLTBEGPACSRPAPVTFPTARLRFVVRNREKRALLYKRNLAQVR 1665

RESULT 11
US-10-697-263-2
; Sequence 2, Application US/10697263
; Publication No. US20040063142A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00927-CIP-DIV2
; CURRENT APPLICATION NUMBER: US/10/697,263
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-697-263-2
```

```
Query Match 20.4%; Score 8423; DB 15; Length 1665;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6356 PPSMOVTIEDVQAOQTGGTAQFEAIIIEGDPQPSVTWKDSVQLVDSTRLSQQOEGTTYSLV 6415
DB 53 PPSMOVTIEDVQAOQTGGTAQFEAIIIEGDPQPSVTWKDSVQLVDSTRLSQQOEGTTYSLV 112
QY 6416 LHRVASKDAGVYVTCIAQNTGGQVLCCKAELLVLGGDNEPDSEKQSHRRKLSHSFYVEKEETG 6475
DB 113 LHRVASKDAGVYVTCIAQNTGGQVLCCKAELLVLGGDNEPDSEKQSHRRKLSHSFYVEKEETG 172
QY 6476 RGVFGFVRVQHKGNKILCAAKFTPLRSRTAQAYRERDILAAALSHPLVTGLLDQFETRK 6535
DB 173 RGVFGFVRVQHKGNKILCAAKFTPLRSRTAQAYRERDILAAALSHPLVTGLLDQFETRK 232
QY 6536 TLILILELCSSEELLDRLYRKGVTAEVKYIQQLVGLHSHGVHLHLDIKPSNIUM 6595
DB 233 TLILILELCSSEELLDRLYRKGVTAEVKYIQQLVGLHSHGVHLHLDIKPSNIUM 292
QY 6596 VHPAREDIKICDFGPAQNITPAELQFSQYSGSPFVSPRIIQONPVSEASDIWAMGVISYL 6655
DB 293 VHPAREDIKICDFGPAQNITPAELQFSQYSGSPFVSPRIIQONPVSEASDIWAMGVISYL 352
QY 6656 SLTCSPPAGSDRATLNLVLEGRVSWSSPMAAHLSEDAKOPFIKATLQAPQAPSAQOC 6715
DB 353 SLTCSPPAGSDRATLNLVLEGRVSWSSPMAAHLSEDAKOPFIKATLQAPQAPSAQOC 412
QY 6716 LSHPWFLKSMAPAEAHFTINTKQLFLLARSWRQSLMSYKSLVWRSIPELLRGPDPSPS 6775
DB 413 LSHPWFLKSMAPAEAHFTINTKQLFLLARSWRQSLMSYKSLVWRSIPELLRGPDPSPS 472
QY 6776 LGVARHLCRDTGGSSSSSSSSSDNELAPPARAKSLPPSPVTHSPLLRGFLRPSASLPEE 6835
DB 473 LGVARHLCRDTGGSSSSSSSSSDNELAPPARAKSLPPSPVTHSPLLRGFLRPSASLPEE 532
QY 6836 AEASERSTEAPAPPASPEGAGPPAAQCGVPRHSVIRSLFYHQAGSPBEHGALAPGSRHRP 6895
DB 533 AEASERSTEAPAPPASPEGAGPPAAQCGVPRHSVIRSLFYHQAGSPBEHGALAPGSRHRP 592
QY 6896 ARRRHLLKGGYTAGALPGLRPLMBHRVLEBEAAREEQATLLAKAPSPETALRLPASGTH 6955
```

```
Db 593 ARRHLLKGGYTAGALPGLRPLMBHRVLEEEAAREEQATLLAKAPSFETALRLPASGTH 652
Qy 6956 LAPGSHSLSHDPSPTPRPSSACCEAQRLLPSAPGGAPIRDMGHPOGSKQLPSTGGHPG 7015
Db 653 LAPGSHSLSHDPSPTPRPSSACCEAQRLLPSAPGGAPIRDMGHPOGSKQLPSTGGHPG 712
Qy 7016 TAQPERPSPDSWPGQAPAFCHPKQSGAPQEGCSPHPAVACPDPGSPFGSCKEAPLPVSS 7075
Db 713 TAQPERPSPDSWPGQAPAFCHPKQSGAPQEGCSPHPAVACPDPGSPFGSCKEAPLPVSS 772
Qy 7076 PFLGQOPAPPAPAKASPLDLSKMGPGDLSLPGRPKPGPCSSPGSASQASSQVSSLRVGS 7135
Db 773 PFLGQOPAPPAPAKASPLDLSKMGPGDLSLPGRPKPGPCSSPGSASQASSQVSSLRVGS 832
Qy 7136 SOVGTEPGSLDAEGWTQAEADLSSTPTLQRPQCAWTKRFSLGGRGYAGVAGTGTA 7195
Db 833 SOVGTEPGSLDAEGWTQAEADLSSTPTLQRPQCAWTKRFSLGGRGYAGVAGTGTA 892
Qy 7196 FCGDAGGMLGQGPMMARIATAVYSQSEEEQEEARAESEEQEEARAEPLPQVSARVP 7255
Db 893 FCGDAGGMLGQGPMMARIATAVYSQSEEEQEEARAESEEQEEARAEPLPQVSARVP 952
Qy 7256 EVGRAPTRSSPEPTWEDIGQVSLVQIRDLSDAEADTISLDISEVDPAYNLSDLYDI 7315
Db 953 EVGRAPTRSSPEPTWEDIGQVSLVQIRDLSDAEADTISLDISEVDPAYNLSDLYDI 1012
Qy 7316 KYLPPEFMIFRKVPKSAPEPPSWAEELAEFPPTWPGELGPHAGLETTESEVDV 7375
Db 1013 KYLPPEFMIFRKVPKSAPEPPSWAEELAEFPPTWPGELGPHAGLETTESEVDV 1072
Qy 7376 ALLAAAVGRKRKWSPPSRSLPHFPGRLPLDEPAELGLRERVKASVBIHSLILKGRPEG 7435
Db 1073 ALLAAAVGRKRKWSPPSRSLPHFPGRLPLDEPAELGLRERVKASVBIHSLILKGRPEG 1132
Qy 7436 LEKEGPPRKCLASFRLSGLKSWDRAPTFLRELSDETAVLQSVTLACQVSAQPAQAAT 7495
Db 1133 LEKEGPPRKCLASFRLSGLKSWDRAPTFLRELSDETAVLQSVTLACQVSAQPAQAAT 1192
Qy 7496 WSKDAPLESSRVLISATLKNFOLLTILVVAEDLGVYTCVSNALGTVTITGVLRKAE 7555
Db 1193 WSKDAPLESSRVLISATLKNFOLLTILVVAEDLGVYTCVSNALGTVTITGVLRKAE 1252
Qy 7556 RPSSSPCPDIEGVYADGVLVWKPVEYGPVTYIVQCSLEGSSWTLASDIFDCCLTSK 7615
Db 1253 RPSSSPCPDIEGVYADGVLVWKPVEYGPVTYIVQCSLEGSSWTLASDIFDCCLTSK 1312
Qy 7616 LSRGTYTFTACVSKAGWGPVSSPEOVLLGGPSHLASEESQGRSAQPLPSTKTFAQ 7675
Db 1313 LSRGTYTFTACVSKAGWGPVSSPEOVLLGGPSHLASEESQGRSAQPLPSTKTFAQ 1372
Qy 7676 TQIQRGFSVVRQCWEKASGRALAAKIIIPYHPKDKTAVLREYALKGLRHPHLAQLHAAY 7735
Db 1373 TQIQRGFSVVRQCWEKASGRALAAKIIIPYHPKDKTAVLREYALKGLRHPHLAQLHAAY 1432
Qy 7736 LSPRHLVILILELCSGPELLPCLAEARASYSESEVKDYLWQMLSATQYLHNQIHLLDLRSE 7795
Db 1433 LSPRHLVILILELCSGPELLPCLAEARASYSESEVKDYLWQMLSATQYLHNQIHLLDLRSE 1492
Qy 7796 NMIIITEYNLLKVVDLIGNAQSLSOEKLPSDKFKDYLETWAPELLEGGQAVPQTDIWAIGV 7855
Db 1493 NMIIITEYNLLKVVDLIGNAQSLSOEKLPSDKFKDYLETWAPELLEGGQAVPQTDIWAIGV 1552
Qy 7856 TAFIMLSAEYPVSSSEGDRLQRLKGLVRLSRVYAGLSGGAVAFRLSTLCAQWGRPCA 7915
Db 1553 TAFIMLSAEYPVSSSEGDRLQRLKGLVRLSRVYAGLSGGAVAFRLSTLCAQWGRPCA 1612
Qy 7916 SSCLOCPWLTEBGPACSRPAPVTPTARLRFVVRNREKRALLLYKRHNLAQVR 7968
Db 1613 SSCLOCPWLTEBGPACSRPAPVTPTARLRFVVRNREKRALLLYKRHNLAQVR 1665
```

RESULT 12

US-10-921-168-2

```
; Sequence 2, Application US/10921168
; Publication No. US20050003446A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO00927-CIP-DIV3
; CURRENT APPLICATION NUMBER: US/10/921,168
; CURRENT FILING DATE: 2004-08-19
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-921-168-2
```

```
Query Match 20.4%; Score 8423; DB 16; Length 1665;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 6356 PPSQVITIEDVOAQTGCTAQPEAIIEGDPQSPVTVYKDSVOLVDSTRLSQQOEGTYSVLV 6415
Db 53 PPSQVITIEDVOAQTGCTAQPEAIIEGDPQSPVTVYKDSVOLVDSTRLSQQOEGTYSVLV 112
Qy 6416 LRHVASKDAGVYTCCLAQNTGGQVLCKAEALLVGLGDNEDPDSEKQSHRRKLHSFYEVKEIG 6475
Db 113 LRHVASKDAGVYTCCLAQNTGGQVLCKAEALLVGLGDNEDPDSEKQSHRRKLHSFYEVKEIG 172
Qy 6476 RGVEFGVKRVQHKNGKILCAAKFTPLRSRTAAQYRERDILAAALSHPLVTGLLQDFETRK 6535
Db 173 RGVEFGVKRVQHKNGKILCAAKFTPLRSRTAAQYRERDILAAALSHPLVTGLLQDFETRK 232
Qy 6536 TLILILELCSSEELLDRLYRQVVTAEVKYIIOQLVEGLHYLHSHGVHLHDIKPSNLLM 6595
Db 233 TLILILELCSSEELLDRLYRQVVTAEVKYIIOQLVEGLHYLHSHGVHLHDIKPSNLLM 292
Qy 6596 VHPAREDIKICDFGAQNIITPAELQFQYQSPFVSPPEIIQONPVSEASDIWAMGVISYL 6655
Db 293 VHPAREDIKICDFGAQNIITPAELQFQYQSPFVSPPEIIQONPVSEASDIWAMGVISYL 352
Qy 6656 SLTCSPPAGESDRATLLNVLEGRVSWSSPMAAHLSEDAKDFIKATLQAPQAPSAQAOC 6715
Db 353 SLTCSPPAGESDRATLLNVLEGRVSWSSPMAAHLSEDAKDFIKATLQAPQAPSAQAOC 412
Qy 6716 LSHPWFLKSPAEAEAHFINTKQLKFLARSWOBSLMSYSKILVMRSTPELLRGPDPSPS 6775
Db 413 LSHPWFLKSPAEAEAHFINTKQLKFLARSWOBSLMSYSKILVMRSTPELLRGPDPSPS 472
Qy 6776 LGVARHLCRDTGGSSSSSSSDNDELAPFARAKSLPPSPVTHSPVTHPLHPRGFLRPSASLPEE 6835
Db 473 LGVARHLCRDTGGSSSSSSSDNDELAPFARAKSLPPSPVTHSPVTHPLHPRGFLRPSASLPEE 532
Qy 6836 AEASERSTEAPAPASPEGAGPPAAQOCVPHSVIRSLFYHQAGESPEHGALAPGSRHP 6895
Db 533 AEASERSTEAPAPASPEGAGPPAAQOCVPHSVIRSLFYHQAGESPEHGALAPGSRHP 592
Qy 6896 ARRHLLKGGYTAGALPGLRPLMEHRVLEEEAAREEQATLLAKAPSFETALRLPASGTH 6955
Db 593 ARRHLLKGGYTAGALPGLRPLMEHRVLEEEAAREEQATLLAKAPSFETALRLPASGTH 652
Qy 6956 LAPGSHSLSHDPSPTPRPSSACCEAQRLLPSAPGGAPIRDMGHPOGSKQLPSTGGHPG 7015
Db 653 LAPGSHSLSHDPSPTPRPSSACCEAQRLLPSAPGGAPIRDMGHPOGSKQLPSTGGHPG 712
Qy 7016 TAQPERPSPDSWPGQAPAFCHPKQSGAPQEGCSPHPAVACPDPGSPFGSCKEAPLPVSS 7075
```

```
Db 713 TAQPERSPDSWGGQAPPCHPKQSAPOEGGSPHPAVAPCPGSPFPGSCKEADLVPS 772
Qy 7076 PFLGQFQAPPAPAKASPLDSSXMGPCDLSLPGKPKPGPCSPGASQASSQVSSLRVGS 7135
Db 773 PFLGQFQAPPAPAKASPLDSSXMGPCDLSLPGKPKPGPCSPGASQASSQVSSLRVGS 832
Qy 7136 SOVGTEPGSLDAEGWTOAEADLSSTPTLQRPQOATMRKESLGRGGYAGVAGYGFPA 7195
Db 833 SOVGTEPGSLDAEGWTOAEADLSSTPTLQRPQOATMRKESLGRGGYAGVAGYGFPA 892
Qy 7196 FCGDAGGMLGQCPMWARAWAYSQSEEBQEBARASQSEEBQEBARASPLQVQARVPV 7255
Db 893 FCGDAGGMLGQCPMWARAWAYSQSEEBQEBARASQSEEBQEBARASPLQVQARVPV 952
Qy 7256 EVGRAPTRSSPEPTWEDIGQVSLVQIRDLSDGAADTISLDSIVDPAYLNLSDLYDI 7315
Db 953 EVGRAPTRSSPEPTWEDIGQVSLVQIRDLSDGAADTISLDSIVDPAYLNLSDLYDI 1012
Qy 7316 KYLPFERMI FRKVPKSAQPEPPMAEBELAFPEPTWPGELGPHAGLEITSESDVD 7375
Db 1013 KYLPFERMI FRKVPKSAQPEPPMAEBELAFPEPTWPGELGPHAGLEITSESDVD 1072
Qy 7376 ALLAAAVCRKRWSSPSRSLFHPGRHLPLDPAELGLRERVKASVEHISRILKGRPEG 7435
Db 1073 ALLAAAVCRKRWSSPSRSLFHPGRHLPLDPAELGLRERVKASVEHISRILKGRPEG 1132
Qy 7436 LEKEGPPRKKGLASFRSLGKSWDRAPTFRLSDTETVVGQSVTLACQVSAQPAQAAT 7495
Db 1133 LEKEGPPRKKGLASFRSLGKSWDRAPTFRLSDTETVVGQSVTLACQVSAQPAQAAT 1192
Qy 7496 WSKGAPLESSRVLISATLKNFQLTLVVAEDLGVTCSVSNALGTVTTTGVLRKAE 7555
Db 1193 WSKGAPLESSRVLISATLKNFQLTLVVAEDLGVTCSVSNALGTVTTTGVLRKAE 1252
Qy 7556 RPSSPCPDIGVYADGVLLWKPVSYPVTYIVQCSLEGGSWTTLASDIFDCCLYTSK 7615
Db 1253 RPSSPCPDIGVYADGVLLWKPVSYPVTYIVQCSLEGGSWTTLASDIFDCCLYTSK 1312
Qy 7616 LSRGCTYTRTACVSKAGMPYSSPSEQVLLGGPSHLSAEEESQGRSAQPLPSTKTFAPQ 7675
Db 1313 LSRGCTYTRTACVSKAGMPYSSPSEQVLLGGPSHLSAEEESQGRSAQPLPSTKTFAPQ 1372
Qy 7676 TOIQGRFSVVRQCKEASGRALAAKIIPYHPKDKTAVLREYALKGLRHPLHAQHAAY 7735
Db 1373 TOIQGRFSVVRQCKEASGRALAAKIIPYHPKDKTAVLREYALKGLRHPLHAQHAAY 1432
Qy 7736 LSPRHLVILELCSGPELLPCLAEARASYESSEVKDYLWQMLSATQYLHNQHLHLDLRSE 7795
Db 1433 LSPRHLVILELCSGPELLPCLAEARASYESSEVKDYLWQMLSATQYLHNQHLHLDLRSE 1492
Qy 7796 NMIIITEYNLLKVVDLGNASQSEKVLPSDKFYLETWAPBELLBGGQAVPQTDIWAIGV 7855
Db 1493 NMIIITEYNLLKVVDLGNASQSEKVLPSDKFYLETWAPBELLBGGQAVPQTDIWAIGV 1552
Qy 7856 TAFIMLSAEYPVSSGARDLQRLKGLVRLSRCVAGLGGGAVAFRLSTLCAQWGRPCA 7915
Db 1553 TAFIMLSAEYPVSSGARDLQRLKGLVRLSRCVAGLGGGAVAFRLSTLCAQWGRPCA 1612
Qy 7916 SSCLOCPMLTEGPACSRPAPVTFPTARLVFVRNREKRRALLYKRNLAQVR 7968
Db 1613 SSCLOCPMLTEGPACSRPAPVTFPTARLVFVRNREKRRALLYKRNLAQVR 1665
```

RESULT 13

US-10-415-011-22

```
; Sequence 22, Application US/10415011
; Publication No. US20040053394A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: CHAWLA, Narinder K.
```

```
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: DING, Li
; APPLICANT: TANG, Y. Tom
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: NGUYEN, Dannie B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LU, Yan
; APPLICANT: YUE, Henry
; APPLICANT: BUREFORD, Neil
; APPLICANT: BANDMAN, Olga
; APPLICANT: TRIBOLEY, Catherine M.
; APPLICANT: LAL, Preeti G.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: KHAN, Farrah A.
; APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0262 USN
; CURRENT APPLICATION NUMBER: US/10/415,011
; PRIOR FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/47728
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,410
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/244,068
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/245,708
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/247,672
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/249,565
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,730
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/250,807
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040053394A1 7638121CDI
US-10-415-011-22
```

```
Query Match 20.4%; Score 9410; DB 15; Length 1665;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1610; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 6356 PPSMQVTIEDVQAQTGGTAQFEAIIIEGDPQPSVTHYKDSVOLVDSVTRLSQQQEGTYSIV 6415
Db 53 PPSMQVTIEDVQAQTGGTAQFEAIIIEGDPQPSVTHYKDSVOLVDSVTRLSQQQEGTYSIV 112
Qy 6416 LRHVASKDAGVYVTCIAQNTGGQVLCKAELLVLGGDNEPDSEKSHRRKLSHFYVKEEIG 6475
Db 113 LRHVASKDAGVYVTCIAQNTGGQVLCKAELLVLGGDNEPDSEKSHRRKLSHFYVKEEIG 172
Qy 6476 RGVFGFVKRVQHKGNKILCAAKFIPLRSRTRAQYRERDILAAALSHPLVTGLLDQFETRK 6535
Db 173 RGVFGFVKRVQHKGNKILCAAKFIPLRSRTRAQYRERDILAAALSHPLVTGLLDQFETRK 232
Qy 6536 TLILILELCSSEELLDRLYRKGVTAEVKYIQQLVGSLHYLHSHGVHLHLDIKPSNIIIM 6595
```


Db 233 TLILILELCSBELDLRLYRKGVVTEAEVKYIQQLVEGLHYLHSHGVHLHLDIKPSNIML 292
Qy 6596 VHPAREDIKICDFGAQNTTTPAELQFSQGSPEFVSPEIIQONPVSEASDIWAMGVISYL 6655
Db 293 VHPAREDIKICDFGAQNTTTPAELQFSQGSPEFVSPEIIQONPVSEASDIWAMGVISYL 352
Qy 6656 SLTSSPPAGESDRATLLNVLEGRYSWSSPMAHLSEDAKOFIKATLQAPQARPSAAQC 6715
Db 353 SLTSSPPAGESDRATLLNVLEGRYSWSSPMAHLSEDAKOFIKATLQAPQARPSAAQC 412
Qy 6716 LSHPWFLKSMPEABEHAHINTKQLKFLARSQRSLMSYKSLVWRSIPPELLRGPDSPS 6775
Db 413 LSHPWFLKSMPEABEHAHINTKQLKFLARSQRSLMSYKSLVWRSIPPELLRGPDSPS 472
Qy 6776 LGVAHRLCHDTCGSSSSSSSSSSSDELAPPARAKSLPPSVTHSPLLHPRGFLRPSASLPEE 6835
Db 473 LGVAHRLCHDTCGSSSSSSSSSSSDELAPPARAKSLPPSVTHSPLLHPRGFLRPSASLPEE 532
Qy 6836 ABASERSTEAPAPPASPEGAGPPAAQCGCVPRHSVIRSLFYHQAGBSPHGHALAPGSRHP 6895
Db 533 ABASERSTEAPAPPASPEGAGPPAAQCGCVPRHSVIRSLFYHQAGBSPHGHALAPGSRHP 592
Qy 6896 ARRHLLKGGYTAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSPETALRLPASGTH 6955
Db 593 ARRHLLKGGYTAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSPETALRLPASGTH 652
Qy 6956 LAPGHSLSLEHDSPTPRPSSACGEAORLPAPSGGAPIRDMGHPQGSKOLPSTGGHPG 7015
Db 653 LAPGHSLSLEHDSPTPRPSSACGEAORLPAPSGGAPIRDMGHPQGSKOLPSTGGHPG 712
Qy 7016 TAOQPERPSPDWGQAPFCHPKQGSAPQEGCSPPHAPVACPFGPGSPFPGSCKEAPLVSS 7075
Db 713 TAOQPERPSPDWGQAPFCHPKQGSAPQEGCSPPHAPVACPFGPGSPFPGSCKEAPLVSS 772
Qy 7076 PFLGQOQAPPAPAKASPLDLSKMGCDISLPGRPKPGCSSPGSASQASSQVSSLRVS 7135
Db 773 PFLGQOQAPPAPAKASPLDLSKMGCDISLPGRPKPGCSSPGSASQASSQVSSLRVS 832
Qy 7136 SOVGTEPGSLDAEGWTQAEADLSSTPTLQRPQEQATMRKFSLGGCGYAGVAGYGTFA 7195
Db 833 SOVGTEPGSLDAEGWTQAEADLSSTPTLQRPQEQATMRKFSLGGCGYAGVAGYGTFA 892
Qy 7196 FGADAGMLGQGPMMARIATAVVSQSEEEQEARAESQSEEQEARAESPLPQVSARVP 7255
Db 893 FGADAGMLGQGPMMARIATAVVSQSEEEQEARAESQSEEQEARAESPLPQVSARVP 952
Qy 7256 EVGRAPTRSSPEPTPWEDIGQVSLVQIRDLSCDAEAADTISLDISEVDPAYNLISDLYDI 7315
Db 953 EVGRAPTRSSPEPTPWEDIGQVSLVQIRDLSCDAEAADTISLDISEVDPAYNLISDLYDI 1012
Qy 7316 KYLPPEFMI FRKVPKSAQEPSPMAEELAEFPPTWPGELGPHAGLEITSESDVD 7375
Db 1013 KYLPPEFMI FRKVPKSAQEPSPMAEELAEFPPTWPGELGPHAGLEITSESDVD 1072
Qy 7376 ALLAEAAVGRKRKWSPPSRSLPHFPGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEG 7435
Db 1073 ALLAEAAVGRKRKWSPPSRSLPHFPGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEG 1132
Qy 7436 LEKEGPRRKPCGLASFRISGLKSWDRAPTFIRELSDETVVLQSVTLACQVSAQAPAAQAT 7495
Db 1133 LEKEGPRRKPCGLASFRISGLKSWDRAPTFIRELSDETVVLQSVTLACQVSAQAPAAQAT 1192
Qy 7496 WSKGAPLESSESRVLISATLKNFOLLITLVVAEDLGVTYTVSVSNALGTVTITGVLRKAE 7555
Db 1193 WSKGAPLESSESRVLISATLKNFOLLITLVVAEDLGVTYTVSVSNALGTVTITGVLRKAE 1252
Qy 7556 RPSSSPCPDIEGVYADGVLLVWKPVSYPVTYIVQCSLEGSWTTLASDIFDCCYLTSK 7615
Db 1253 RPSSSPCPDIEGVYADGVLLVWKPVSYPVTYIVQCSLEGSWTTLASDIFDCCYLTSK 1312
Qy 7616 LSRGCTYTRTACVSKACMGVSPSPSEOVLLGGPSHLASEESQGRSAQPLPSTKTFAFO 7675
Db 1313 LSRGCTYTRTACVSKACMGVSPSPSEOVLLGGPSHLASEESQGRSAQPLPSTKTFAFO 1372

Qy 7676 TOIQGRFSVVRQCEKASGRALAAKIIPYHPKDKTAVLREYALKGLRHPHLAOLHAAY 7735
Db 1373 TOIQGRFSVVRQCEKASGRALAAKIIPYHPKDKTAVLREYALKGLRHPHLAOLHAAY 1432
Qy 7736 LSPRHLVLILBELCGPELLPCLABRASYSSEVKDYLMQMLSATQYLNQHLHLDLRSE 7795
Db 1433 LSPRHLVLILBELCGPELLPCLABRASYSSEVKDYLMQMLSATQYLNQHLHLDLRSE 1492
Qy 7796 NMIIITEYNLLKVDLGNASQSLQSKVLPSPDKFKDYLETMAPELLEGGQAVPQTDIWAIGV 7855
Db 1493 NMIIITEYNLLKVDLGNASQSLQSKVLPSPDKFKDYLETMAPELLEGGQAVPQTDIWAIGV 1552
Qy 7856 TAFIMLSAEYPVSSSEGARDLQRLKGLVRLSRVAGLSGGAVAFPLRSTLCAQPWGRPCA 7915
Db 1553 TAFIMLSAEYPVSSSEGARDLQRLKGLVRLSRVAGLSGGAVAFPLRSTLCAQPWGRPCA 1612
Qy 7916 SSCLQCPWLTEBGPACSRPAPVPTPTARLRFVVRNREKRALLYKRHNLAQVR 7968
Db 1613 SSCLQCPWLTEBGPACSRPAPVPTPTARLRFVVRNREKRALLYKRHNLAQVR 1665

RESULT 14

US-10-182-243-46
; Sequence 46, Application US/10182243
; Publication No. US20040048310A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY D.
; APPLICANT: WHYTE, DAVID
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; APPLICANT: MARTINEZ, RICARDO
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE
; TITLE OF INVENTION: ENZYMS
; FILE REFERENCE: 038602/1366
; CURRENT APPLICATION NUMBER: US/10/182,243
; CURRENT FILING DATE: 2003-07-07
; PRIOR APPLICATION NUMBER: PCT/US01/02337
; PRIOR FILING DATE: 2001-01-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 1618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-182-243-46

Query Match 20.4%; Score 8407; DB 15; Length 1618;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

Qy 6357 PSMQVTTIEDVQAQGTGTAQFEAIIIEGDPQPSVTWKDSVOLVDSTRLSQOEGTYSVL 6416
Db 1 PSMQVTTIEDVQAQGTGTAQFEAIIIEGDPQPSVTWKDSVOLVDSTRLSQOEGTYSVL 60
Qy 6417 RHVASKDAGVYTCLAQNTGGQVLCKAEALLVLG-----GNEPDSEKQSHRRKLHSFEV 6470
Db 61 RHVASKDAGVYTCLAQNTGGQVLCKAEALLVLGASHLSIGDNEPDSEKQSHRRKLHSFEV 120
Qy 6471 KEETIGRVGFGVRKQVHGKNIKCAAKFIPLRSTRQAAYRERDIIAALSHPLVTGLDQ 6530
Db 121 KEETIGRVGFGVRKQVHGKNIKCAAKFIPLRSTRQAAYRERDIIAALSHPLVTGLDQ 180
Qy 6531 FETRKTLLILELCSBELDLRLYRKGVVTEAEVKYIQQLVEGLHYLHSHGVHLHLDIKP 6590
Db 181 FETRKTLLILELCSBELDLRLYRKGVVTEAEVKYIQQLVEGLHYLHSHGVHLHLDIKP 240
Qy 6591 SNILMVHPAREDIKICDFGAQNTTTPAELQFSQGSPEFVSPEIIQONPVSEASDIWAMG 6650
Db 241 SNILMVHPAREDIKICDFGAQNTTTPAELQFSQGSPEFVSPEIIQONPVSEASDIWAMG 300
Qy 6651 VISYLSLTCSPPAGESDRATLLNVLEGRYSWSSPMAHLSEDAKOFIKATLQAPQARP 6710

QY	6839	SERSTEAPAPPASPEGAGPPAAQGCVPRHVSIRSLFYHQGESPBHGALAPGSRHRPARR	6898
Db	481	SERSTEAPAPPASPEGAGPPAAQGCVPRHVSIRSLFYHQGESPBHGALAPGSRHRPARR	540
QY	6899	RHLKGGYIAGALPGLREPMHRVLEBEAAREEQATLLAKAPSPETALRLPASGTHLAP	6958
Db	541	RHLKGGYIAGALPGLREPMHRVLEBEAAREEQATLLAKAPSPETALRLPASGTHLAP	600
QY	6959	GHSLSLEHDSPTTPRSSACGEAORLPSAPSGGAPIRDMGHPQSGKOLPSTGGHPGTAQ	7018
Db	601	GHSLSLEHDSPTTPRSSACGEAORLPSAPSGGAPIRDMGHPQSGKOLPSTGGHPGTAQ	660
QY	7019	PERPSPDPSWGPAPFCHPKQGSAPQEGCSHPAVAPCPGSPFGSCKEAPLVPSSPFL	7078
Db	661	PERPSPDPSWGPAPFCHPKQGSAPQEGCSHPAVAPCPGSPFGSCKEAPLVPSSPFL	720
QY	7079	GQOQAPPAPAKASPLDLSKMGPGDISLPGRPKPGPCSPGSSAQSSQVSLRVGSSQV	7138
Db	721	GQOQAPPAPAKASPLDLSKMGPGDISLPGRPKPGPCSPGSSAQSSQVSLRVGSSQV	780
QY	7139	GTEPGPSLDAEGWTOEAEDLSSTPTLORPOQATMRKFSLGGRGYAGVAGYGTFAFGG	7198
Db	781	GTEPGPSLDAEGWTOEAEDLSSTPTLORPOQATMRKFSLGGRGYAGVAGYGTFAFGG	840
QY	7199	DAGGMLGQGPMMARIAMAVSQSEEBEQAERAEQSEEQEAREASPLQVSARVPVEVG	7258
Db	841	DAGGMLGQGPMMARIAMAVSQSEEBEQAERAEQSEEQEAREASPLQVSARVPVEVG	900
QY	7259	RAPTRSSPEPTWEDIGQVSLVQIRDLSGDABAADTISLDISEVDPAVNLNLDYDIKYL	7318
Db	901	RAPTRSSPEPTWEDIGQVSLVQIRDLSGDABAADTISLDISEVDPAVNLNLDYDIKYL	960
QY	7319	PPEFMI FRKVPKSAQPEPPSPMAEBELAEFPPTWMPGELGPHAGLEITEESDVEDALL	7378
Db	961	PPEFMI FRKVPKSAQPEPPSPMAEBELAEFPPTWMPGELGPHAGLEITEESDVEDALL	1020
QY	7379	ABAAVGRKRWSSPSRSLFHPFGRHLPLDEPAELGLRERVKASVEHISRILKGRPEGLEK	7438
Db	1021	ABAAVGRKRWSSPSRSLFHPFGRHLPLDEPAELGLRERVKASVEHISRILKGRPEGLEK	1080
QY	7439	EGPRPKKGLASFRISGLKSWDRAPFLREISDETVVLGOSVTLACQVSAQPAQAATWSK	7498
Db	1081	EGPRPKKGLASFRISGLKSWDRAPFLREISDETVVLGOSVTLACQVSAQPAQAATWSK	1140
QY	7499	DCAPLESSSRVLI SATLKNFOLLTLVVVAEDLVYTCVSNALGTVTITGVLRKAERPS	7558
Db	1141	DCAPLESSSRVLI SATLKNFOLLTLVVVAEDLVYTCVSNALGTVTITGVLRKAERPS	1200
QY	7559	SSPCPDIGEVVADGVLLVWKPVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSLSR	7618
Db	1201	SSPCPDIGEVVADGVLLVWKPVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSLSR	1260
QY	7619	GGTYTFRCTACVKAGMGYPSPSEOVTLGGPSHLASEESQGRSAQPLPSTKTFAFQOI	7678
Db	1261	GGTYTFRCTACVKAGMGYPSPSEOVTLGGPSHLASEESQGRSAQPLPSTKTFAFQOI	1320
QY	7679	QRGFSVVRQWEKASGRALAAKIIPYHPKOKTAVLREYEALKGLRHPHLAQLHAAVLSLSP	7738
Db	1321	QRGFSVVRQWEKASGRALAAKIIPYHPKOKTAVLREYEALKGLRHPHLAQLHAAVLSLSP	1380
QY	7739	RHLVLIILELCSGPELLLPCLAERASYESSEVKDYLMQMLSATQYLNQHIHLHLDRSENMI	7798
Db	1381	RHLVLIILELCSGPELLLPCLAERASYESSEVKDYLMQMLSATQYLNQHIHLHLDRSENMI	1440
QY	7799	ITEYNLLKVVLDLGNQSLSOEKLVPSPDKFQYLETMAPELLEGGAVPOTDIWAIGVTAF	7858
Db	1441	ITEYNLLKVVLDLGNQSLSOEKLVPSPDKFQYLETMAPELLEGGAVPOTDIWAIGVTAF	1500
QY	7859	IMLSAEYPVSESGARDLQRLRKGILVRLSRCYAGLSGGAVAFRLRSTLCAQPWGRPCASSC	7918
Db	1501	IMLSAEYPVSESGARDLQRLRKGILVRLSRCYAGLSGGAVAFRLRSTLCAQPWGRPCASSC	1560

QY	7919	LQCPWLTEEGPACSRPAPVTPTPTARLRFVFNREKRALLYKRHNLAQVR	7968
Db	1561	LQCPWLTEEGPACSRPAPVTPTPTARLRFVFNREKRALLYKRHNLAQVR	1610

Search completed: March 18, 2005, 18:05:29
Job time : 554 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2005, 17:32:23 ; Search time 126 Seconds
(without alignments)
4720.665 Million cell updates/sec

Title: US-10-077-130-5
Perfect score: 41273
Sequence: 1 MDQFSGCAPFLTRPKAFV.....RNREKRALLYRHNLAQVR 7968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8423	20.4	1665	4	US-09-858-664A-2
2	8423	20.4	1665	4	US-10-274-978-2
3	8423	20.4	1665	4	US-10-697-263-2
4	1173	2.8	846	4	US-09-858-664A-3
5	1173	2.8	846	4	US-10-274-978-4
6	1173	2.8	846	4	US-10-697-263-4
7	1074	2.6	4391	4	US-10-006-011A-2
8	1025	2.5	1953	4	US-09-917-254-92
9	909.5	2.2	549	4	US-09-858-664A-5
10	909.5	2.2	549	4	US-10-274-978-6
11	909.5	2.2	549	4	US-08-826-267-2
12	892.5	2.2	2860	2	US-08-826-267-2
13	592.5	1.4	414	4	US-09-858-664A-13
14	592.5	1.4	414	4	US-10-274-978-14
15	592.5	1.4	414	4	US-10-697-263-14
16	572.5	1.4	1709	4	US-09-949-016-10503
17	542.5	1.3	615	4	US-09-949-016-11320
18	531	1.3	279	4	US-09-858-664A-4
19	531	1.3	279	4	US-10-274-978-5
20	531	1.3	279	4	US-10-697-263-5
21	521	1.3	298	4	US-09-858-664A-17
22	521	1.3	298	4	US-10-274-978-18
23	521	1.3	298	4	US-10-697-263-18
24	511.5	1.2	508	4	US-09-858-664A-18
25	511.5	1.2	508	4	US-10-274-978-19
26	511.5	1.2	508	4	US-10-697-263-19
27	504	1.2	250	4	US-09-858-664A-6

28	504	1.2	250	4	US-10-274-978-7	Sequence 7, Appl1
29	504	1.2	250	4	US-10-697-263-7	Sequence 7, Appl1
30	497.5	1.2	596	4	US-09-797-039-8	Sequence 8, Appl1
31	495	1.2	969	4	US-09-949-016-8059	Sequence 8059, Ap
32	489	1.2	448	2	US-09-159-385-2	Sequence 2, Appl1
33	489	1.2	448	3	US-09-186-277-2	Sequence 2, Appl1
34	476.5	1.2	11877	3	US-09-105-537-6	Sequence 6, Appl1
35	472.5	1.1	260	2	US-07-857-224B-23	Sequence 23, Appl
36	471	1.1	1431	4	US-09-538-092-1198	Sequence 1198, Ap
37	470	1.1	502	4	US-09-949-016-7837	Sequence 7837, Ap
38	466.5	1.1	454	2	US-09-159-385-1	Sequence 1, Appl1
39	466.5	1.1	454	3	US-09-186-277-1	Sequence 1, Appl1
40	466.5	1.1	454	4	US-09-949-016-6066	Sequence 6066, Ap
41	466	1.1	274	4	US-09-858-664A-14	Sequence 14, Appl
42	466	1.1	274	4	US-10-274-978-15	Sequence 15, Appl
43	466	1.1	274	4	US-10-697-263-15	Sequence 15, Appl
44	465	1.1	358	4	US-09-230-896C-29	Sequence 29, Appl
45	464.5	1.1	1050	3	US-09-428-711A-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-858-664A-2
; Sequence 2, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-2

Query Match	20.4%	Score 8423;	DB 4;	Length 1665;
Best Local Similarity	99.9%	Pred. NO. 0;		
Matches 1612;	Conservative	0;	Mismatches	1;
			Indels	0;
			Gaps	0;
Qy	6356	PPSMQVTTIEDVQAQTGGTAQFAEIIIEGDPQPSVTWKDSVQLVDSTRLSQQOEGTYSILV	6415	
Db	53	PPSMQVTTIEDVQAQTGGTAQFAEIIIEGDPQPSVTWKDSVQLVDSTRLSQQOEGTYSILV	112	
Qy	6416	LRHVASKDAGVYVTCIAQNTGQVLCKAELLVGGNEDPSEKQSHRRKLHSPYVEKEIG	6475	
Db	113	LRHVASKDAGVYVTCIAQNTGQVLCKAELLVGGNEDPSEKQSHRRKLHSPYVEKEIG	172	
Qy	6476	RGVFGVXRVQHKGNKILCAAKFIPLRSRTAQAYRERDIIAALSHPLVTGLDQFETRK	6535	
Db	173	RGVFGVXRVQHKGNKILCAAKFIPLRSRTAQAYRERDIIAALSHPLVTGLDQFETRK	232	
Qy	6536	TLILILECSSEELLDRLYRGVGVTEAEVKYIQQLVEGLHSHGVHLHDIKPSNIIIM	6595	
Db	233	TLILILECSSEELLDRLYRGVGVTEAEVKYIQQLVEGLHSHGVHLHDIKPSNIIIM	292	
Qy	6596	VHPAREDIKIDFGFAQNTITPAELQFSQYSGPFSPIIQONPVSEASDIWAMGVISYL	6655	
Db	233	VHPAREDIKIDFGFAQNTITPAELQFSQYSGPFSPIIQONPVSEASDIWAMGVISYL	352	
Qy	6656	SLTCSPPAGESDRATLLNVLEGRVSWSSPMAAHLSEDAKDFIKATLQAPQAPSAQC	6715	
Db	353	SLTCSPPAGESDRATLLNVLEGRVSWSSPMAAHLSEDAKDFIKATLQAPQAPSAQC	412	

[illegible]

Qy	6956	LAPGHSHLEHDSPTPRPSSBCEAQRLPSAPSGGAPIDRMGHPOGSKOLPSTGCHPG	7015
Db	653	LAPGHSHLEHDSPTPRPSSBCEAQRLPSAPSGGAPIDRMGHPOGSKOLPSTGCHPG	712
Qy	7016	TAQPERPSPDWGQAPFCHPKQGSAPQEGCSPHPAVAPCPPGSPFGSCKEAPLPVSS	7075
Db	713	TAQPERPSPDWGQAPFCHPKQGSAPQEGCSPHPAVAPCPPGSPFGSCKEAPLPVSS	772
Qy	7076	PFLGQOPAPPAPAKASPLDLSKMGPDGDSLPRPKPGPCSSPGSASQASSQSLSRVGS	7135
Db	773	PFLGQOPAPPAPAKASPLDLSKMGPDGDSLPRPKPGPCSSPGSASQASSQSLSRVGS	832
Qy	7136	SOVGTPEGPSLDAEGWTOEADLSDSTPTLORPOEQATMRKFSLGRGGVAGVAGYTGA	7195
Db	833	SOVGTPEGPSLDAEGWTOEADLSDSTPTLORPOEQATMRKFSLGRGGVAGVAGYTGA	892
Qy	7196	FGGDAGMLGQOPMWARLAWAVSOSBEEBEEAARASQSEBQOQARAEPLPQVARSVP	7255
Db	893	FGGDAGMLGQOPMWARLAWAVSOSBEEBEEAARASQSEBQOQARAEPLPQVARSVP	952
Qy	7256	EVGRAPTRSSPEPTPWEDIGQVSLVQIRDLSGDAEAAATISLDISEVDPAYLNLSDLYDI	7315
Db	953	EVGRAPTRSSPEPTPWEDIGQVSLVQIRDLSGDAEAAATISLDISEVDPAYLNLSDLYDI	1012
Qy	7316	KYLDFEFMIFRKVPKSAQPEPSPMAEBEELAEFPEPTWMPGELGPHAGLEITEESDND	7375
Db	1013	KYLDFEFMIFRKVPKSAQPEPSPMAEBEELAEFPEPTWMPGELGPHAGLEITEESDND	1072
Qy	7376	ALLAEAAVGRKRWSSPSRSLFHPGRHLPLDEPAELGLRERVKASVEHISRLILKGRPEG	7435
Db	1073	ALLAEAAVGRKRWSSPSRSLFHPGRHLPLDEPAELGLRERVKASVEHISRLILKGRPEG	1132
Qy	7436	LEKGGPRKPGELASFRLSGLKSWDRAPTFLRELSDETVVLGQSVTLACQVSAQPAQAAT	7495
Db	1133	LEKGGPRKPGELASFRLSGLKSWDRAPTFLRELSDETVVLGQSVTLACQVSAQPAQAAT	1192
Qy	7496	WSDGAPLESSSRVLIISATLKNFOLLITLVVAEDLGVYTCSVSNALGTWTTTGVLRKAE	7555
Db	1193	WSDGAPLESSSRVLIISATLKNFOLLITLVVAEDLGVYTCSVSNALGTWTTTGVLRKAE	1252
Qy	7556	RPSSSPCPDIGEVADGVLVWKPEVSEYGPVYTVIQCSLEGGSWTTILASDIFDCCYLTSK	7615
Db	1253	RPSSSPCPDIGEVADGVLVWKPEVSEYGPVYTVIQCSLEGGSWTTILASDIFDCCYLTSK	1312
Qy	7616	LSRGGTTFRTACVSKAGMGYPSPSQVLLGGPSHLASEEBSQGRSAQPLPSTKTFAPQ	7675
Db	1313	LSRGGTTFRTACVSKAGMGYPSPSQVLLGGPSHLASEEBSQGRSAQPLPSTKTFAPQ	1372
Qy	7676	TQIQGRFVSVRQCEWKASGRALAAKIIIPHPKDKTAVLREYALKGLRHPLHQAOLHAAY	7735
Db	1373	TQIQGRFVSVRQCEWKASGRALAAKIIIPHPKDKTAVLREYALKGLRHPLHQAOLHAAY	1432
Qy	7736	LSPRHLVLIILELCGPELLPCIAERASYSSEVKDYLMQMLSATQYLHNOHILHDLRSE	7795
Db	1433	LSPRHLVLIILELCGPELLPCIAERASYSSEVKDYLMQMLSATQYLHNOHILHDLRSE	1492
Qy	7796	NMIITEYNLLKVDPLGNAQSLSQEKVLPDKFKDYLETMAPELLEGGQAVPOTDIIWAIGV	7855
Db	1493	NMIITEYNLLKVDPLGNAQSLSQEKVLPDKFKDYLETMAPELLEGGQAVPOTDIIWAIGV	1552
Qy	7856	TAFIMLSAEYPVSSSEGARDLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQWPGRPCA	7915
Db	1553	TAFIMLSAEYPVSSSEGARDLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQWPGRPCA	1612
Qy	7916	SSCLOCPWLTEEGPACSRPAPVTPPTTARLVRVFNREKRALLIYKRHNLAQVR	7968
Db	1613	SSCLOCPWLTEEGPACSRPAPVTPPTTARLVRVFNREKRALLIYKRHNLAQVR	1665

RESULT 4

US-09-858-664A-3

Sequence 3, Application US/09858664A

; Patent No. 6482624

```

; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-858-664A-3

```

Query Match 2.8%; Score 1173; DB 4; Length 846;
Best Local Similarity 34.9%; Pred. No. 1.4e-54;
Matches 313; Conservative 121; Mismatches 316; Indels 146; Gaps 29;

Qy	6357	PSHQVITIEDVQAGTGTGAQPEALIEGDPOBSWTWYKDSQVLVDSTRLSQOQEGTTSVLV 6416
Db	1	PFRESIMEDVEVGAGETARFAVVVEGKPLFDIMWYKDEVLTTESHSVFYBENECSLV 60
Qy	6417	RHVASDAGVYTCLAQNTGGOVLCKAELLVLGDN---EPDSEKQSHR-RKLHSFYEVK 6471
Db	61	LSTGAQDQGGVYTCTAQNLAGVSKCAELAVHSAQTAWEVEGVGEDDHRRRLUSDYDTH 120
Qy	6472	EETGRGVGFVKRVQHKNKILCAAKEFIPLRSTRAQAYRERDILAAALSHPLVTGLDOF 6531
Db	121	QEIGRGAFSYLRRIVERSSGLEFAAKFIPSOAKPKASARREARLLARLOHDCVLYPHEAF 180
Qy	6532	ETRKTLILILELCSBELDLRYKRGVVTAEAVKVYIQOLVEGLHYLVHSHGVHLHDIKPS 6591
Db	181	ERRRGHVIWTEL--TEELLERARKPTVCESITRAYMRQVLGEGIHQSHVLHLHDVKE 239
Qy	6592	NIILMVHPA--REDIKICDFGFAQNTIPAEILOFSQVSGSPFVSPETIIQQNPVSEASDIWAM 6649
Db	240	NLLWMDGAGEQVVRICDFGNQELTPGPOCYGTPEFVAPEIVNQSPVSGVTDIWPV 299
Qy	6650	GVISYLSLTCSPPAGESDRATILNLYEGRVSWSPMAHLSSEDADF--IKATLQRAPOA 6708
Db	300	GVVAFCLTGISPFVGENDRITLMLNRYNVAFEETFLLSREARGFLIKVLQV--DRL 357
Qy	6709	RPSAAOCLSHMPLKMPABEAHFINTOKFLLRKRWQORSUMSYKSLVMRSPPELLR 6768
Db	358	RPTABETLHPMPKTOAKGAE--VSTDHLKFLFLRRRWQORSQISYKCHLVLRPIPELLR 414
Qy	6769	GPDSFSLGVARHLCRDTCGSSSSSSSSDNEL-----APPARAK-SLPPSPVTH 6816
Db	415	APPERVWVTMPRR--PPSPGGLSSSDSEEELEELPSVPRPIQPEFSGRVSULTDPTED 473
Qy	6817	SPLLHPRGFLRPSASLPEBAEASERSTEAPAPASPPEGAGPAAAOCCVPRHSVIR----- 6871
Db	474	EALGTETGNAATMDWQEOGRAPSOQOEAPSPALSPQOEPA--GASPRGELRGSSA 532
Qy	6872	-----SLFYHQAG--ESPBGHALAPG-----SRHPARRHLHLK 6903
Db	533	ESALPRAGPRELGRGLHKAASVELPQRRSPGPGATFLARGGLGEYEAORLOALQRLLR 592
Qy	6904	GGYIACALPGLREPLMEH-----RVLEEEAAREEQATL---LAKAPSPETALR 6948
Db	593	GGPEDGVSLGRLGPLLESJGGRRDRPMARAASSAPHHQPLENRGLQKSSFSQGSA 652
Qy	6949	LPASGTHLAPGHSHSLHSDSPSTPR----PSSEACGEAQRLPSPASGCGAIPRDMGHQGS 7004
Db	653	EP-RGRHRRAGAPLEIPVARLGRRLQESFSLSEAO--FSSPA-----RPSAP 700
Qy	7005	KQLPSTGGHPGTAQBERPSDPSWGOPAPFCHPKQGSAPQEGCSBHPAVACPPOGFPFG 7064
Db	701	K--PST---PKAEASATPDSAPOPAP--QPADKAPPEPPEPVRASKAP-----PPQ 749


```
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-697-263-6

Query Match          2.2%; Score 909.5; DB 4; Length 549;
Best Local Similarity 38.3%; Pred. No. 1.1e-40;
Matches 204; Conservative 88; Mismatches 181; Indels 59; Gaps 12;

QY 6274 TGTSEAPVPPRPVLPOLLHEGPEQPEAIARAQEWTVIRMEGAAMPAGGT-----GE 6326
Db 25 TVKSSKSPSPPEPQLLEHGTLE-EAPAMLDKPDIVVVGQ--PASVTVTFNHVEAQ 81
QY 6327 LMDVHSHVVRITORTTYQAID-----THTAR----- 6355
Db 82 VVMRSRCGALLEARAGVYELSQPDDQYCLRICRVSRDRMGALTCARNRHGTQTCVTL 141
QY 6356 ---PPSMQVTIEDVQAOTGGTAQFEALIEGDPQPSVTWYKDSVOLVDSTRLSSQQOEGTT 6411
Db 142 ELAEAPRFRESIMEDVEVGAGETARFAVVVEGKPLDIMWYKDEVLLTSSSHVSFVYEENE 201
QY 6412 YSLVLRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGGDN----BPDSEKQSHR-RKLHS 6466
Db 202 CSLVVLSTGAQDGGVYTCTAQNLAGESVCKAELAVHSAQTAMEVGVGVEDDHRGRLSD 261
QY 6467 FYEVKEETGRGVGVKRVQHKNGKILCAAKPIPLRSTRQAQYRERDILALSHPLVTG 6526
Db 262 FYDIHQEIRGAFSLRIVERSSGLEFAAKFIPSOAKPKASARREARLLARLQHDVCVLY 321
QY 6527 LLDQFETRTKLILILELCSSEBLLDLRYKGVVTEAEVKVYIQQLAVEGLHYLHSHGVHL 6586
Db 322 FHEAFERRGLVITELC-TEELLERIAKPTVCESEIRAYMRQVLEGHLYHSHGVHL 380
QY 6587 DIKPSNLMVHPA--REDIKICDFGAQNIITPAELQFQSGSPFVSPSEIIQQNPVSRAS 6644
Db 381 DVKPNLLAVDGAAGEQQVRCDFGNAQBELTGPGEQYCYGTPEFVAPEIVNQSPVSGVT 440
QY 6645 DIWAMGVLSYLSLTCSPPAGESDRATLLNVLEGRVSWSSPMAHLSDAKDF-IKATLQ 6703
Db 441 DIMPVGVVAFPLCLTGISPFVGENDRITLLMNRYNVAFEETTFLSREARGLIKVLVQ 500
QY 6704 RAPQARPSAAOCLSHPWFLKSPAPAEAHFINTKOLKFLLRSRWORSUMSYK 6755
Db 501 --DLRPTAETLEHPWFKTQAKGAE---VSTDHLKFLSRRRWORSQISYK 547

RESULT 12
US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070
; GENERAL INFORMATION:
; APPLICANT: Streuli, Michel
; TITLE OF INVENTION: No. 5994070el TRIO Molecules and Uses Related Thereto
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826.267
; FILING DATE: 1997
; CLASSIFICATION: 800
```

QY 6269 SPPVTCTSEAPVPRVPQPLLHSGPEQPEPAIARAQEWTPVIRMEGANWFGAGTGELL 6328
Db 2423 YRKSREGLSN-----KVSVKLLNP----- 2441
QY 6329 WDVHSHVRETTQRTYTYQAIDHTARPSPSMOVTIEDVQAOTGGTAQPEAIEGDPQPSV 6388
Db 2442 -----NYIDV----- 2480
QY 6389 TWYKDSVOLVD-----STRLSQQQGGTYSVLVRHVASKDAGVYTCCLAQNTGQVLCXAB 6443
Db 2481 TWKGPEHNTLNNDGHYSISYDLGEAT---LKIIVGVTTEDDGIYTCIAVNDMGSSASSAS 2537
QY 6444 LLVLGGDNEPSEKSHRKLHSFYEVKEEIGRGVFGVGRVQVHKGKILCAAKETPLRS 6503
Db 2538 LRVLFGP---MDGIMVTWKDNFDSFYSEVAELGRGFSVVKCDQGTGRAVATKFNKKL 2595
QY 6504 RTRAQYRERDILAALSHPLVTGLLDQFETRTKLILILELSCSSEBLLDLRYKGVVTEAE 6563
Db 2596 MKRDQVTHELGILQSLQHPLLVGLLDTFETPTSYILVLEMAQDQRLDCVVRWGSLSLTKG 2655
QY 6564 VKVYIQQLVEGLYHSHGVHLDTKPSNLMVHP-AREDIKICDFGFAQNTTPAELQFS 6622
Db 2656 ITRAHGEVLEAVRYLHNCRIAHLDLKPENILVDESIAKPTIKLADFGDAVQNLNTYYIHQ 2715
QY 6623 QYGSPEFVSPELQONPVSEASDIWAMGVISYLSLTCSSPPAGESDRATLNLVLEGRVSW 6682
Db 2716 LLGNPEFAAPEILGNPVSLSDTWSVGLTYVLLSGVSPFLDSDVEETCLNICRLDPSF 2775
QY 6683 SSPMAAHLSEDAKOFIKATLQAPQAPSAOCLSHPMFLKSMPEAEAHFINTKOLKFL 6742
Db 2776 PDDYFKGVSKAKKEFVCLLDQEDPAKRSALALQEQW-LQAGNGRSTGVLDTSRLTSFI 2834
QY 6743 ARSRQW---RSLMSYKSLVMRSIP 6764
Db 2835 ERKHQNDVVRPIRSIKNFLQSLRP 2859

RESULT 13

US-09-858-664A-13
; Sequence 13, Application US/09858664A
; Patent No. 6482624

; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP

; CURRENT APPLICATION NUMBER: US/09/858,664A

; PRIOR FILING DATE: 2001-05-17

; PRIOR FILING DATE: 2001-05-17

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 414

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-858-664A-13

Query Match 1.4%; Score 592.5; DB 4; Length 414;
Best Local Similarity 34.5%; Pred. No. 7.6e-24;
Matches 145; Conservative 69; Mismatches 191; Indels 15; Gaps 6;

QY 6356 PPSMQVTIEDVQAOTGGTAQPEAIEGDPQPSVTWKDSVOLVD-----STRLSQQQGT 6410
Db 1 PPEFVIPLEVTCTGETVWLRCVGRPKASITWKGEHNTLNNDGHYSISYDLGEAT 60

QY 6411 TYSVLVRHVASKDAGVYTCCLAQNTGGVLCXAEALLVGGDNEPSEKSHRKLHSFYEV 6470
Db 61 ---LKIVGVTTEDDGIYTCIAVNDMGSSASSASLRVLGPG--MDGIMVTWKDNFDSFYSE 115

QY 6471 KEETIGRGVFGVGRVQVHKGKILCAAKETPLRSRTRAQYRERDILAALSHPLVTGLLDQ 6530

Db 116 VAEELGRGFSVVKCDQGTGRAVATKFNKKLMKRDQVTHELGILQSLQHPLLVGLLD 175
QY 6531 PETRTKLILILELSCSSEBLLDLRYKGVVTEAEVYVYIQQLVEGLYHSHGVHLDTK 6590
Db 176 FETPTSYILVLEMAQDQRLDCVVRWGSLSLTKGIRAHLEGEVLEAVRYLHNCRIAHLD 235
QY 6591 SNILMVHP-AREDIKICDFGFAQNTTPAELQFSQYGSPEFVSPELQONPVSEASDIWAM 6649
Db 236 ENILVDESIAKPTIKLADFGDAVQNLNTYYIHQLLGNPEFAAPEILGNPVSLSDTWSV 295
QY 6650 GVISYLSLTCSSPPAGESDRATLNLVLEGRVSWSSPMAAHLSEDAKOFIKATLQAPQAR 6709
Db 236 GVLITVLLSGVSPFLDSDVEETCLNICRLDPSFDDYFKGVSKAKKEFVCLLDQEDPAK 355
QY 6710 PSAAOCLSHPMFLKSMPEAEAHFINTKOLKFLARSRWQ---RSLMSYKSLVMRSIP 6766
Db 356 PSALALQEQW-LQAGNGRSTGVLDTSRLTSFIERRKHQNDVVRPIRSIKNFLQSLRP 414

RESULT 14

US-10-274-978-14

; Sequence 14, Application US/10274978

; Patent No. 6670164

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui, et al

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL000927-CIP-DIV

; CURRENT APPLICATION NUMBER: US/10/274,978

; PRIOR FILING DATE: 2002-10-22

; PRIOR FILING DATE: 2001-05-17

; PRIOR FILING DATE: 2001-05-17

; PRIOR FILING DATE: 2000-11-14

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 14

; LENGTH: 414

; TYPE: PRT

; ORGANISM: Human

US-10-274-978-14

Query Match 1.4%; Score 592.5; DB 4; Length 414;
Best Local Similarity 34.5%; Pred. No. 7.6e-24;
Matches 145; Conservative 69; Mismatches 191; Indels 15; Gaps 6;

QY 6356 PPSMQVTIEDVQAOTGGTAQPEAIEGDPQPSVTWKDSVOLVD-----STRLSQQQGT 6410
Db 1 PPEFVIPLEVTCTGETVWLRCVGRPKASITWKGEHNTLNNDGHYSISYDLGEAT 60

QY 6411 TYSVLVRHVASKDAGVYTCCLAQNTGGVLCXAEALLVGGDNEPSEKSHRKLHSFYEV 6470
Db 61 ---LKIVGVTTEDDGIYTCIAVNDMGSSASSASLRVLGPG--MDGIMVTWKDNFDSFYSE 115

QY 6471 KEETIGRGVFGVGRVQVHKGKILCAAKETPLRSRTRAQYRERDILAALSHPLVTGLLDQ 6530
Db 116 VAEELGRGFSVVKCDQGTGRAVATKFNKKLMKRDQVTHELGILQSLQHPLLVGLLD 175

QY 6531 PETRTKLILILELSCSSEBLLDLRYKGVVTEAEVYVYIQQLVEGLYHSHGVHLDTK 6590
Db 176 FETPTSYILVLEMAQDQRLDCVVRWGSLSLTKGIRAHLEGEVLEAVRYLHNCRIAHLD 235

QY 6591 SNILMVHP-AREDIKICDFGFAQNTTPAELQFSQYGSPEFVSPELQONPVSEASDIWAM 6649
Db 236 ENILVDESIAKPTIKLADFGDAVQNLNTYYIHQLLGNPEFAAPEILGNPVSLSDTWSV 295

QY 6650 GVISYLSLTCSSPPAGESDRATLNLVLEGRVSWSSPMAAHLSEDAKOFIKATLQAPQAR 6709
Db 296 GVLITVLLSGVSPFLDSDVEETCLNICRLDPSFDDYFKGVSKAKKEFVCLLDQEDPAK 355

QY 6710 PSAAOCLSHPMFLKSMPEAEAHFINTKOLKFLARSRWQ---RSLMSYKSLVMRSIP 6766

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2005, 17:25:27 ; Search time 575 Seconds
(without alignments)
5359.490 Million cell updates/sec

Title: US-10-077-130-5

Perfect score: 41273

Sequence: 1 MDQPFSCAPFLTRPKAFV.....RNREKRALLYKRNHLAQVR 7968

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp1980e:*

2: geneseqp1990e:*

3: geneseqp2000e:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41273	100.0	7968	6	ABG76187 Human ser
2	32134	77.9	6620	7	ADJ70485 Human hea
3	13710	33.2	2630	6	ABG76186 Human ser
4	13528	32.8	2596	4	AAB30569 A splice
5	10519.5	25.5	4691	5	ABP70084 Human NOV
6	10506	25.5	4675	5	ABP70085 Human NOV
7	9834.5	23.8	2328	6	ABP58227 Human cel
8	8423	20.4	1665	5	AAO15372 Human myo
9	8410	20.4	1665	5	AAE24151 Human kin
10	8407	20.4	1618	4	AAB85504 Human pro
11	8403	20.4	1610	4	AAB30568 A full le
12	7073	17.1	1351	4	AAB30567 Amino aci
13	6619.5	16.0	1596	7	ADJ69186 Human hea
14	4533	11.0	871	4	AAB30570 A splice
15	4529	11.0	871	5	AAE16274 Human kin
16	3878.5	9.4	31267	6	ABG74786 Human RGS
17	3580.5	8.7	26926	4	AAU05396 Human tit
18	3577.5	8.7	26926	8	ADQ17316 Human sof
19	3427	8.3	34350	8	ADQ89964 Antagonis
20	2489	6.0	6642	8	ADN22360 Bacteri
21	2447.5	5.9	548	4	AAB30571 A full le
22	2415.5	5.9	548	4	AAB30572 A full le
23	2360	5.7	4412	3	AAV53666 Sequence
24	2328.5	5.6	3267	7	ADQ99075 Human KPP
25	2324	5.6	3208	7	ADG47676 Human NOV

26	2324	5.6	3208	8	ADJ78946	Adj78946 Human NOV
27	2319	5.6	3268	7	ADB79959	Adb79959 Human kin
28	2270	5.5	3252	7	ADG47674	Adg47674 Human NOV
29	2270	5.5	3252	8	ADJ78944	Adj78944 Human NOV
30	2203	5.3	3262	7	ADB79961	Adb79961 Mouse ser
31	2192.5	5.3	5614	8	ADR66062	Adr66062 Human pro
32	2192.5	5.3	5614	8	ADR66404	Adr66404 Human pro
33	2135	5.2	2380	5	AAE19160	Aae19160 Human kin
34	2127	5.2	2286	4	AAB65635	Aab65635 Novel pro
35	2127	5.2	2286	8	ADJ29242	Adj29242 Human MAR
36	2102	5.1	5635	5	ABP60991	Abp60991 Novel hum
37	2085.5	5.1	3186	7	ADG47672	Adg47672 Human NOV
38	2085.5	5.1	3186	8	ADJ78942	Adj78942 Human NOV
39	2085	5.1	416	7	ADM05704	Adm05704 Human pro
40	2072.5	5.0	5636	7	ADJ70089	Adj70089 Human hea
41	2072.5	5.0	5636	7	ADJ81337	Adj81337 Human hem
42	2072.5	5.0	5636	8	ADK60205	Adk60205 Angiogene
43	2072.5	5.0	5636	8	ADK60506	Adk60506 Angiogene
44	2072.5	5.0	5636	8	ADP73129	Adp73129 Angiogene
45	2067	5.0	2231	7	ADB79962	Adb79962 Human KIA

ALIGNMENTS

RESULT 1

ABG76187
ID ABG76187 standard; protein; 7968 AA.

XX AC ABG76187;

XX DT 09-MAY-2003 (first entry)

XX DE Human serine/threonine or protein kinase 12599.

XX KW Human; enzyme; serine/threonine kinase; protein kinase; 12599;

XX KW cardiovascular disease; heart failure; myocardial infarction;

XX KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma; immunogen;

XX KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;

XX KW haemolytic anaemia; cellular proliferative disorder; cancer;

XX KW protein kinase disorder; autoimmune disorder; diabetes mellitus;

XX KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;

XX KW multiple sclerosis.

XX OS Homo sapiens.

XX PN US2002168742-A1.

XX PD 14-NOV-2002.

XX PF 15-FEB-2002; 2002US-00077130.

XX PR 15-FEB-2001; 2001US-0269201P.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Kapeller-Libermann R, Acton SL;

XX DR WPI; 2003-298729/29.

XX DR N-PSDB; ABX11642.

XX PT Novel isolated human protein kinase, designated 59079 or 12599

XX PT polypeptide, useful as diagnostic and therapeutic agents for preventing

XX PT cardiovascular diseases, proliferative disorders, and protein kinase

XX disorders.

XX PS Claim 8; Page 84-104; 119pp; English.

XX The invention relates to an isolated human serine/threonine or protein

XX kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule

XX comprising at least 85% identity to the nucleic acids appearing as

XX ABX11641 and ABX11642 or their complement, a naturally occurring variant

XX of the kinases or their fragments. Also included are a non-human host

cell containing the nucleic acids, an antibody specific for the proteins, identifying a compound which binds to the kinase (by contacting the kinase or a cell expressing the kinase with a test compound and determining whether the kinase binds to the test compound) and modulating the activity of kinase using the identified compound. The kinases and their encoding nucleic acids are useful as diagnostic and therapeutic agents for preventing a disease or condition associated with an aberrant or unwanted 59079 or 12599 activity in a subject, including cardiovascular diseases such as heart failure, and myocardial infarction; disorders involving blood vessels such as atherosclerosis, and Kaposi's sarcoma; blood platelets disorder such as thrombocytopaenia, and leukaemia; Hodgkin's disease, hemolytic anaemia; cellular proliferative disorders such as cancer; and protein kinase disorders such as autoimmune disorders, diabetes mellitus, psoriasis, inflammatory bowel disease, rheumatoid arthritis, and multiple sclerosis (many examples of diseases and disorders are included in the specification). The kinases, their encoding nucleic acids and antibodies are useful in screening assays, detection assays (e.g. forensic biology), and predictive medicine (e.g. diagnostic assays, prognostic assays, and monitoring clinical trials and pharmacogenomics). The kinases and their encoding nucleic acids are useful as query sequences to perform a search against public databases to identify other family members or related sequences. The present sequence represents kinase 12599

XX Sequence 7968 AA;

Query Match	100.0%;	Score 41273;	DB 6;	Length 7968;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 7968;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MDQFSGAPRLTRPKAFVSVGKDATLSQIVGNPTFQVSWKDXDQPVTAGARFLAQ	60	
DB	1	MDQFSGAPRLTRPKAFVSVGKDATLSQIVGNPTFQVSWKDXDQPVTAGARFLAQ	60	
QY	61	DGDLRLTLIDLALGDSGYVCRANAI GEAPAAVGLQVDAEAAAEQAPFLRLPTIR	120	
DB	61	DGDLRLTLIDLALGDSGYVCRANAI GEAPAAVGLQVDAEAAAEQAPFLRLPTIR	120	
QY	121	VREGSEATFRCRVGSPPAPVSWSKDGRRLGPDGPRVVRVEELGEASALIRARPRDGG	180	
DB	121	VREGSEATFRCRVGSPPAPVSWSKDGRRLGPDGPRVVRVEELGEASALIRARPRDGG	180	
QY	181	TYEVAENPLGAASAAAALWDSDAADTASRPGTSTAALLAHLQRRREAMEAGAPASPP	240	
DB	181	TYEVAENPLGAASAAAALWDSDAADTASRPGTSTAALLAHLQRRREAMEAGAPASPP	240	
QY	241	STGTRTCTVTGKHARLSCYVTGEPKPTVWKDQQLVTEGRRHVYVEDAQNFLKILF	300	
DB	241	STGTRTCTVTGKHARLSCYVTGEPKPTVWKDQQLVTEGRRHVYVEDAQNFLKILF	300	
QY	301	CKQSDRGLYTCTASNLVGQTVSSVLVVRPEAPVPFKRLQDLVREKESATFLCEVPQPS	360	
DB	301	CKQSDRGLYTCTASNLVGQTVSSVLVVRPEAPVPFKRLQDLVREKESATFLCEVPQPS	360	
QY	361	TEAAMFKETRLWASAKYIGIEEGTERLTVRNVSADDDAVYICETPEGSRVLAELAVQG	420	
DB	361	TEAAMFKETRLWASAKYIGIEEGTERLTVRNVSADDDAVYICETPEGSRVLAELAVQG	420	
QY	421	NLLRLPKRTAVRVGDTAMFCVELAVPGVPHMLRNQEVVAGGRVAISAGTRHTLTIS	480	
DB	421	NLLRLPKRTAVRVGDTAMFCVELAVPGVPHMLRNQEVVAGGRVAISAGTRHTLTIS	480	
QY	481	QCCLDEVGVQAFMAGDCQTSFRFCVSAAPRKPLOPPDVPVVKARMESSVILSWSPPHGE	540	
DB	481	QCCLDEVGVQAFMAGDCQTSFRFCVSAAPRKPLOPPDVPVVKARMESSVILSWSPPHGE	540	
QY	541	RPVTIDGVLVEKKLGTWTWIRCHAEAWATPELTADVABEGNFPQFRVSALNSFGQSPY	600	
DB	541	RPVTIDGVLVEKKLGTWTWIRCHAEAWATPELTADVABEGNFPQFRVSALNSFGQSPY	600	
QY	601	LEFPCTVHLAKPLAVRTPKAVQAVEGGEVTFSDLTIVASAGFWFLDGOALKASSVYEH	660	
DB	601	LEFPCTVHLAKPLAVRTPKAVQAVEGGEVTFSDLTIVASAGFWFLDGOALKASSVYEH	660	

QY	661	CDRTRHTLTIREVPASLHGAOLKFVANGIESSIRMEVRAAPGLTANKPPAAAAAEVLARL	720	
DB	661	CDRTRHTLTIREVPASLHGAOLKFVANGIESSIRMEVRAAPGLTANKPPAAAAAEVLARL	720	
QY	721	HEEAQLLAELSQAQAAAVTWLKDGRTLSPGPKYEVQASAGRRVLLVRDVARDDAGLYECVS	780	
DB	721	HEEAQLLAELSQAQAAAVTWLKDGRTLSPGPKYEVQASAGRRVLLVRDVARDDAGLYECVS	780	
QY	781	RGGRVIAQLSVQGLARFLHDKMAGSCVDVAVAGGPAQPECETSEAHVHVHWTYKDGMELGHS	840	
DB	781	RGGRVIAQLSVQGLARFLHDKMAGSCVDVAVAGGPAQPECETSEAHVHVHWTYKDGMELGHS	840	
QY	841	GERFLOEDVGRHRLVAATVTRQDEGTVSVCRVGDSDVDRLRVSEPKVVFVFAKEQLARRKL	900	
DB	841	GERFLOEDVGRHRLVAATVTRQDEGTVSVCRVGDSDVDRLRVSEPKVVFVFAKEQLARRKL	900	
QY	901	QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMPEATGCTRRLLVVQAGQADAGEY	960	
DB	901	QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMPEATGCTRRLLVVQAGQADAGEY	960	
QY	961	SCEAGGQRLSPHLVDVKEPKVVFVFAKDOVAHSEVQAEAGANNATLSCEVAQAQAEVVMYKDGK	1020	
DB	961	SCEAGGQRLSPHLVDVKEPKVVFVFAKDOVAHSEVQAEAGANNATLSCEVAQAQAEVVMYKDGK	1020	
QY	1021	KLSSSLKVHVEAKGCRRLRVVQAGKTDAGDYSCEARGQVSRFLRHITTEPKMFAKEQSV	1080	
DB	1021	KLSSSLKVHVEAKGCRRLRVVQAGKTDAGDYSCEARGQVSRFLRHITTEPKMFAKEQSV	1080	
QY	1081	HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVGMVEYKGTCTRRLLVLPQAGKAD	1140	
DB	1081	HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVGMVEYKGTCTRRLLVLPQAGKAD	1140	
QY	1141	AGEYSCAGGQVSRFLHITTEPKGVFAKEQSVHNEVQAEAGTAMLSCEVAQAQTEVTWY	1200	
DB	1141	AGEYSCAGGQVSRFLHITTEPKGVFAKEQSVHNEVQAEAGTAMLSCEVAQAQTEVTWY	1200	
QY	1201	KDGKLSSSSKVMEVKGCTRRLLVVQVQKADAGEYSCAGGQVSRFQHLHITEPKAVPAK	1260	
DB	1201	KDGKLSSSSKVMEVKGCTRRLLVVQVQKADAGEYSCAGGQVSRFQHLHITEPKAVPAK	1260	
QY	1261	EQLVHNEVTRTEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVIRIEAAGCMRQLVVOQA	1320	
DB	1261	EQLVHNEVTRTEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVIRIEAAGCMRQLVVOQA	1320	
QY	1321	QADAGEYTCAGGQRLSPHLVDVSEPKAVFAKEQLAHRKVQAEAGATATLSCEVAQAQTE	1380	
DB	1321	QADAGEYTCAGGQRLSPHLVDVSEPKAVFAKEQLAHRKVQAEAGATATLSCEVAQAQTE	1380	
QY	1381	VTWYKDGKLSSSSKVMEVKGCTRRLLVVQVQACQADTGEYSCAGGQRLSPSLDVAEPKV	1440	
DB	1381	VTWYKDGKLSSSSKVMEVKGCTRRLLVVQVQACQADTGEYSCAGGQRLSPSLDVAEPKV	1440	
QY	1441	VFAKEQPVHREVQQAQAGATATLSCEVAQAQTEVTWYKDGKLSSSSKVMEVAVGCTRRLLV	1500	
DB	1441	VFAKEQPVHREVQQAQAGATATLSCEVAQAQTEVTWYKDGKLSSSSKVMEVAVGCTRRLLV	1500	
QY	1501	VQQAQADAGEYSCAGGQRLSPHLHVAEPKAVFAKEQPAQASREVQAEAGTSATLSCEVAQ	1560	
DB	1501	VQQAQADAGEYSCAGGQRLSPHLHVAEPKAVFAKEQPAQASREVQAEAGTSATLSCEVAQ	1560	
QY	1561	AQTEVTWYKDGKLSSSSKVMEVAVGCTRRLLVVQAEQADAGEYSCKAGQRLSFHLHVA	1620	
DB	1561	AQTEVTWYKDGKLSSSSKVMEVAVGCTRRLLVVQAEQADAGEYSCKAGQRLSFHLHVA	1620	
QY	1621	EPKVVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVREAVGCT	1680	
DB	1621	EPKVVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVREAVGCT	1680	
QY	1681	RRLVVQQAQADAGEYSCAGGQRLSPHLHVAEIPQISERPCCRREPLVVVKEHEDITLTA	1740	
DB	1681	RRLVVQQAQADAGEYSCAGGQRLSPHLHVAEIPQISERPCCRREPLVVVKEHEDITLTA	1740	

QY 1741 TLATPSAAVTWIKDGVETRRSKHETASQGDHHTLTVHGAQVLDIAIYSCRVBAGQDP 1800
DB 1741 TLATPSAAVTWIKDGVETRRSKHETASQGDHHTLTVHGAQVLDIAIYSCRVBAGQDP 1800
QY 1801 PVQVEEVAAKFRLLEPVCGLGGTITLACELSPACAEVVRWCGNTQPRVGRFQWVAEG 1860
DB 1801 PVQVEEVAAKFRLLEPVCGLGGTITLACELSPACAEVVRWCGNTQPRVGRFQWVAEG 1860
QY 1861 PVRSUTVLGLRAEDAGEYCESRDDHTSAQLTVSVPRVVKFMSGLSTVVAEBGGATPQC 1920
DB 1861 PVRSUTVLGLRAEDAGEYCESRDDHTSAQLTVSVPRVVKFMSGLSTVVAEBGGATPQC 1920
QY 1921 VVPSDVAVVRPDGALLQPSKPAISQSGASHSLTISDLVLEDAGQITVVEABGSSAA 1980
DB 1921 VVPSDVAVVRPDGALLQPSKPAISQSGASHSLTISDLVLEDAGQITVVEABGSSAA 1980
QY 1981 LRVREAPVLFKKLJFPQTVVEERSSVTLELTPMPPELRWTRNATALAPKKNVEHAEGA 2040
DB 1981 LRVREAPVLFKKLJFPQTVVEERSSVTLELTPMPPELRWTRNATALAPKKNVEHAEGA 2040
QY 2041 RHRLVLHNVGFPADRGFFGCETPDDKTQAKLTVMEMQVRLVRLGLOAVEAREOQTATMEVOL 2100
DB 2041 RHRLVLHNVGFPADRGFFGCETPDDKTQAKLTVMEMQVRLVRLGLOAVEAREOQTATMEVOL 2100
QY 2101 SHADVGSWTRDGLRFQOQPTCHLAVRGPMTLITLSGLRPEDSGLMVFKAEGVHTSARLV 2160
DB 2101 SHADVGSWTRDGLRFQOQPTCHLAVRGPMTLITLSGLRPEDSGLMVFKAEGVHTSARLV 2160
QY 2161 VTLEPVSFSPRLQDVVTTEKEKVTLECELSRNVDRVLKDGVELRAGKTMALAAQACR 2220
DB 2161 VTLEPVSFSPRLQDVVTTEKEKVTLECELSRNVDRVLKDGVELRAGKTMALAAQACR 2220
QY 2221 SLTIYRCBFDAGGVVCDADHAQSSASVKVQGRVTYTLIYRRVLAEDAGEIQFVAENAESR 2280
DB 2221 SLTIYRCBFDAGGVVCDADHAQSSASVKVQGRVTYTLIYRRVLAEDAGEIQFVAENAESR 2280
QY 2281 AOLRVKELPVTILVRLRDKIANEKHGVLECOVSRASQVRWFKGSOELQPKFVLSVD 2340
DB 2281 AOLRVKELPVTILVRLRDKIANEKHGVLECOVSRASQVRWFKGSOELQPKFVLSVD 2340
QY 2341 GLYRLKLIISDVHAEDEDTYTCAGDVKTSQAQFFVEEQSITIVRGLQDVTVMPEPAPWPEC 2400
DB 2341 GLYRLKLIISDVHAEDEDTYTCAGDVKTSQAQFFVEEQSITIVRGLQDVTVMPEPAPWPEC 2400
QY 2401 ETSIPSPRPKWLKGLTKVLQAGNVGLEQBGTVHRLMLRRTCTMTGPHVFTVGRSSSA 2460
DB 2401 ETSIPSPRPKWLKGLTKVLQAGNVGLEQBGTVHRLMLRRTCTMTGPHVFTVGRSSSA 2460
QY 2461 RLWVSDIPVLTPLRPLEKPTGRELOQSVLSCDPRPAPKAVQWYKDDTPLSPSEKFKWGLEG 2520
DB 2461 RLWVSDIPVLTPLRPLEKPTGRELOQSVLSCDPRPAPKAVQWYKDDTPLSPSEKFKWGLEG 2520
QY 2521 QMAELRILRLMPADAGVYRCQAGSAHSTEVTVAREVTVTGLODAEATEEGMASFSCE 2580
DB 2521 QMAELRILRLMPADAGVYRCQAGSAHSTEVTVAREVTVTGLODAEATEEGMASFSCE 2580
QY 2581 LSHDEEVEWSLNGMPLYNDSFPEISHKGRHRTLVLKSTIQRADAGIVRASSLKVSTARS 2640
DB 2581 LSHDEEVEWSLNGMPLYNDSFPEISHKGRHRTLVLKSTIQRADAGIVRASSLKVSTARS 2640
QY 2641 EVRVKPVVFLKALDLSAEERTLALQCEVSDPEAHVVRKDGVLQPSDKYDFLHTAGT 2700
DB 2641 EVRVKPVVFLKALDLSAEERTLALQCEVSDPEAHVVRKDGVLQPSDKYDFLHTAGT 2700
QY 2701 RGLVVDVSPEDAGLYTCHVGSEETRARVRVHDLHVGITKRLKTMVELEGSCSPECVLS 2760
DB 2701 RGLVVDVSPEDAGLYTCHVGSEETRARVRVHDLHVGITKRLKTMVELEGSCSPECVLS 2760
QY 2761 HESASDPAMWTGGKTVGSSSRFQATROGRKYILVVRREAPSDAGEVVFVSRGLTSKASL 2820
DB 2761 HESASDPAMWTGGKTVGSSSRFQATROGRKYILVVRREAPSDAGEVVFVSRGLTSKASL 2820
QY 2821 IVRERPAAIKPLEDQWVAPGEDVELRCELRSAGTFVHMLKDKRAIRKSKYQYDVVCEGTM 2880

DB 2821 IVRERPAAIKPLEDQWVAPGEDVELRCELRSAGTFVHMLKDKRAIRKSKYQYDVVCEGTM 2880
QY 2881 AMLVIRGASLKDAGEYTCVEASKSTASLHVBEKANCFTPEELTNLQVEBKGTAVFTCKTE 2940
DB 2881 AMLVIRGASLKDAGEYTCVEASKSTASLHVBEKANCFTPEELTNLQVEBKGTAVFTCKTE 2940
QY 2941 HPAATVTVWRKGLLELRASGHKQPSQEGTLRLTITISALEKADSDTYTCDIQOASRAQLLV 3000
DB 2941 HPAATVTVWRKGLLELRASGHKQPSQEGTLRLTITISALEKADSDTYTCDIQOASRAQLLV 3000
QY 3001 QGRRVHHIILEDVDVQEGSSATFCRISPNANYEPVHMFDTKTPHANELEIDAQPGGY 3060
DB 3001 QGRRVHHIILEDVDVQEGSSATFCRISPNANYEPVHMFDTKTPHANELEIDAQPGGY 3060
QY 3061 HVLTLRQLALKDSGTIYFEAGDQASAAALRVTEKPSVFSRELTDATITEGEDTLVCETS 3120
DB 3061 HVLTLRQLALKDSGTIYFEAGDQASAAALRVTEKPSVFSRELTDATITEGEDTLVCETS 3120
QY 3121 TCDIPMCWTGDKTLRGSAARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSIVRV 3180
DB 3121 TCDIPMCWTGDKTLRGSAARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSIVRV 3180
QY 3181 HARPVRFQEAUKDLEVLGGATLRCVLSVAAPVKWCYGNVLRPGDKYSLRQEGAMLE 3240
DB 3181 HARPVRFQEAUKDLEVLGGATLRCVLSVAAPVKWCYGNVLRPGDKYSLRQEGAMLE 3240
QY 3241 LVVRNLRFDQSGRYSCSFGDQTTSATLTVTALPAOFIIGKLRNKEATGATATLRCBSLKT 3300
DB 3241 LVVRNLRFDQSGRYSCSFGDQTTSATLTVTALPAOFIIGKLRNKEATGATATLRCBSLKT 3300
QY 3301 APVWRKGSSETLRDGDYCLRDQDGMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPM 3360
DB 3301 APVWRKGSSETLRDGDYCLRDQDGMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPM 3360
QY 3361 PAHFTIGRLRHOESIEGATATLRCELSKAAPVWRKGRSLRDGDHSLRQDGAVALCQIC 3420
DB 3361 PAHFTIGRLRHOESIEGATATLRCELSKAAPVWRKGRSLRDGDHSLRQDGAVALCQIC 3420
QY 3421 GLAVADAGEYSCVCGEERTSATLTVKALPAKFTGELRNEEAVEGATAMLCELSVAPE 3480
DB 3421 GLAVADAGEYSCVCGEERTSATLTVKALPAKFTGELRNEEAVEGATAMLCELSVAPE 3480
QY 3481 WRKGPENLRDGDYILRQEGTCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARF 3540
DB 3481 WRKGPENLRDGDYILRQEGTCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARF 3540
QY 3541 IEDVKNOQAREGATAVLOCELNSAAPVWRKGSSETLRDGDYSLRQDGTKCELOIRGLAM 3600
DB 3541 IEDVKNOQAREGATAVLOCELNSAAPVWRKGSSETLRDGDYSLRQDGTKCELOIRGLAM 3600
QY 3601 ADTGEYSCVCGQERTSAMLTVRALPIKFTGELRNEEATGATAVLRCELSKNAPVEMWKG 3660
DB 3601 ADTGEYSCVCGQERTSAMLTVRALPIKFTGELRNEEATGATAVLRCELSKNAPVEMWKG 3660
QY 3661 HETLRDGDHSLRQDGAARCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKFIEGL 3720
DB 3661 HETLRDGDHSLRQDGAARCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKFIEGL 3720
QY 3721 RNEEATEGDTATLWCELSKAAPVWRKGHETLRDGDHSLRQDGSRCCELOIRGLAVDAG 3780
DB 3721 RNEEATEGDTATLWCELSKAAPVWRKGHETLRDGDHSLRQDGSRCCELOIRGLAVDAG 3780
QY 3781 EYSCVCGQERTSATLTVRALPARFTEDVKNOQAREGATAVLOCELSKAAPVEMWKGSETL 3840
DB 3781 EYSCVCGQERTSATLTVRALPARFTEDVKNOQAREGATAVLOCELSKAAPVEMWKGSETL 3840
QY 3841 RGGDRYSLRQDGTTRCELOIHGSLVADTGEYSCVCGQERTSATLTVRAPQVPREPLQSLQ 3900
DB 3841 RGGDRYSLRQDGTTRCELOIHGSLVADTGEYSCVCGQERTSATLTVRAPQVPREPLQSLQ 3900
QY 3901 AEBGSTATLQCELSSEPTATVWVSKGLOLQANGRRPRELQGTAEVLVLODLOREDTGEYT 3960

Db 3901 ABEGSTATLOCELSEPTATVVMXSGGLQLOANGRRRPRLOQCTAELVLQDLQEDBTGEYT 3960
Qy 3961 CTCSSQATSATLTTAAAPVRFELRELQOEVDVEGTAHLCCELSRAGASVEWRKGSLOLEP 4020
Db 3961 CTCSSQATSATLTTAAAPVRFELRELQOEVDVEGTAHLCCELSRAGASVEWRKGSLOLEP 4020
Qy 4021 CAKQMWQDGAABELLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRILQSLQE 4080
Db 4021 CAKQMWQDGAABELLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRILQSLQE 4080
Qy 4081 TGDIAIRCQLSDAESGAUVQWLKEGVELHAGPKYEMRSQATRELLIHOLEAKDTGEYA 4140
Db 4081 TGDIAIRCQLSDAESGAUVQWLKEGVELHAGPKYEMRSQATRELLIHOLEAKDTGEYA 4140
Qy 4141 CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADDEVEFSCVSRAGATGVQWCLQGLPLQ 4200
Db 4141 CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADDEVEFSCVSRAGATGVQWCLQGLPLQ 4200
Qy 4201 SNEVTEVAVRDRGRIHTLRLKGVTPEDACTVSFHLGNHASSAQLTVRAPEVTILEPLODVQ 4260
Db 4201 SNEVTEVAVRDRGRIHTLRLKGVTPEDACTVSFHLGNHASSAQLTVRAPEVTILEPLODVQ 4260
Qy 4261 LSEGODASFQRLSRASQEARWALGGVPLQANEMNDITVQGTLLHLTLHKVTLEDAGT 4320
Db 4261 LSEGODASFQRLSRASQEARWALGGVPLQANEMNDITVQGTLLHLTLHKVTLEDAGT 4320
Qy 4321 VSFHVGTCSSEAOUKVTAKNTVVRGLENVVALEGEALFECQLOPEVAAHTWLLDDEPV 4380
Db 4321 VSFHVGTCSSEAOUKVTAKNTVVRGLENVVALEGEALFECQLOPEVAAHTWLLDDEPV 4380
Qy 4381 RTSNAEVVPFENGRLHLLLLKNLRPQDSRCVTFLAGDMVTSAFILTVRGMRLILEPLKN 4440
Db 4381 RTSNAEVVPFENGRLHLLLLKNLRPQDSRCVTFLAGDMVTSAFILTVRGMRLILEPLKN 4440
Qy 4441 AAVRAGAQRFTCTLSEAVPVGEASWYINGAAVOPDDSDWTATDAGSHQALLLLSAQPHH 4500
Db 4441 AAVRAGAQRFTCTLSEAVPVGEASWYINGAAVOPDDSDWTATDAGSHQALLLLSAQPHH 4500
Qy 4501 AGEVTFACRDVAVASARLTVLGLPDPEDAEEVVAHSSHVTLSWAAPMSDGGGLCGYRVE 4560
Db 4501 AGEVTFACRDVAVASARLTVLGLPDPEDAEEVVAHSSHVTLSWAAPMSDGGGLCGYRVE 4560
Qy 4561 VKEGATQWRLCHELVPGEPCVVDGLAPGETYRFRVAAGVPVGAGEPVHLPTQTVRLAEP 4620
Db 4561 VKEGATQWRLCHELVPGEPCVVDGLAPGETYRFRVAAGVPVGAGEPVHLPTQTVRLAEP 4620
Qy 4621 KPVPQPSAPESRQVAAGEDVLSLEVVAAEAGEVIVHKGMERIQPGGRFVVSQGRQOML 4680
Db 4621 KPVPQPSAPESRQVAAGEDVLSLEVVAAEAGEVIVHKGMERIQPGGRFVVSQGRQOML 4680
Qy 4681 VIKGFTAEDQGEYHCGLAQSGICPAAATFQVALSPASVDEAPQPSLPPEAAQEGDLHLW 4740
Db 4681 VIKGFTAEDQGEYHCGLAQSGICPAAATFQVALSPASVDEAPQPSLPPEAAQEGDLHLW 4740
Qy 4741 EALARKRMSREPTLDSISELPEEDGRSQRLPQEAEEVAPDLSEGYSTADELARTGDADL 4800
Db 4741 EALARKRMSREPTLDSISELPEEDGRSQRLPQEAEEVAPDLSEGYSTADELARTGDADL 4800
Qy 4801 SHTSDDESAGTSPSLVLYLKAGRPSTSLASKVGAPAAASVKPQQOQEPPLAARVPLG 4860
Db 4801 SHTSDDESAGTSPSLVLYLKAGRPSTSLASKVGAPAAASVKPQQOQEPPLAARVPLG 4860
Qy 4861 DLSTKDLGDPSMDKAANKIQAAFKGYKVRKEMKQEGPMFSHTFGDTQAQVGDALRLFCV 4920
Db 4861 DLSTKDLGDPSMDKAANKIQAAFKGYKVRKEMKQEGPMFSHTFGDTQAQVGDALRLFCV 4920
Qy 4921 VASKADVPAARWLKQGVELTDGRHHHIDQLDGTCSLLIAGLDRADAGCYTCQVSNKFGQV 4980
Db 4921 VASKADVPAARWLKQGVELTDGRHHHIDQLDGTCSLLIAGLDRADAGCYTCQVSNKFGQV 4980
Qy 4981 THSACVVVSGSEASESSGGELDDAFRAARLHRLFRTPSPAEVSDEELFLSADEGPA 5040
Db 4981 THSACVVVSGSEASESSGGELDDAFRAARLHRLFRTPSPAEVSDEELFLSADEGPA 5040

Qy 5041 EPBEPADQWYREDEHFCIRFEALTEARQAVTRFQEMFATLIGIVBEIKLVEQGRPRVEM 5100
Db 5041 EPBEPADQWYREDEHFCIRFEALTEARQAVTRFQEMFATLIGIVBEIKLVEQGRPRVEM 5100
Qy 5101 CISKETPAPVVPPEPLPSLLTSDAAPVFLTELQOEVDQGYPVSFDCVVTGQMPSPVRWF 5160
Db 5101 CISKETPAPVVPPEPLPSLLTSDAAPVFLTELQOEVDQGYPVSFDCVVTGQMPSPVRWF 5160
Qy 5161 KDGKLLBEDDHYMINEDQGGHQLIITAVVPADMVGVVRCCLAENSMGVSSYTKAEULRVDLTS 5220
Db 5161 KDGKLLBEDDHYMINEDQGGHQLIITAVVPADMVGVVRCCLAENSMGVSSYTKAEULRVDLTS 5220
Qy 5221 TDYDTAADATSESSYFSAQGYLSRREQEGTSTTDEGQLPQVVEELRDLQVAPGTRLAKE 5280
Db 5221 TDYDTAADATSESSYFSAQGYLSRREQEGTSTTDEGQLPQVVEELRDLQVAPGTRLAKE 5280
Qy 5281 QLKVKGYPAARLWFKDGOQPLTASAHIRMTGKKILHTLEIISVTRDESQGYAAVISNAMG 5340
Db 5281 QLKVKGYPAARLWFKDGOQPLTASAHIRMTGKKILHTLEIISVTRDESQGYAAVISNAMG 5340
Qy 5341 AAYSSARLLVRGDPDEPEKPADSVHEQLVPPMLERFTPKVKKGSSITTSVKVEGRPVP 5400
Db 5341 AAYSSARLLVRGDPDEPEKPADSVHEQLVPPMLERFTPKVKKGSSITTSVKVEGRPVP 5400
Qy 5401 TVHWRREEABRGVLWIGPDTPGYTVASSAQOHSIIVLLDVGRQHOGTYTCTIASNAAGQALC 5460
Db 5401 TVHWRREEABRGVLWIGPDTPGYTVASSAQOHSIIVLLDVGRQHOGTYTCTIASNAAGQALC 5460
Qy 5461 SASLHVSGLPKVBEOEKVKEALISTFLOGTQOALSQAGLETASPADIGGQKKEPLAAKE 5520
Db 5461 SASLHVSGLPKVBEOEKVKEALISTFLOGTQOALSQAGLETASPADIGGQKKEPLAAKE 5520
Qy 5521 ALGHLISLAEVGTREFLOKLTQITTEMVSAKITQAKLOVPGGDSDEDSKTSASPRHGSR 5580
Db 5521 ALGHLISLAEVGTREFLOKLTQITTEMVSAKITQAKLOVPGGDSDEDSKTSASPRHGSR 5580
Qy 5581 PSSSIOESSSESDGDARGEIFDIYVVTADYLPJGABQDAITTLREGQYVEVLDAAHPLRW 5640
Db 5581 PSSSIOESSSESDGDARGEIFDIYVVTADYLPJGABQDAITTLREGQYVEVLDAAHPLRW 5640
Qy 5641 LVRTKPTKSPSGQWSPAYLDRLKLSPEWGAABEPEPGEAVSDEYKARLSSVIOE 5700
Db 5641 LVRTKPTKSPSGQWSPAYLDRLKLSPEWGAABEPEPGEAVSDEYKARLSSVIOE 5700
Qy 5701 LLSSEQAFVSELOPQSHHLQHLERCPHPTAVAGQKAVIFRNVDRDIGRPHSSFLQELQQ 5760
Db 5701 LLSSEQAFVSELOPQSHHLQHLERCPHPTAVAGQKAVIFRNVDRDIGRPHSSFLQELQQ 5760
Qy 5761 CDTDDDDVAMCFIKNQAAFEQYLEFLVGRVQAESVVVSTAOEFYKKYABEALLAGDPSQP 5820
Db 5761 CDTDDDDVAMCFIKNQAAFEQYLEFLVGRVQAESVVVSTAOEFYKKYABEALLAGDPSQP 5820
Qy 5821 PPPLOHLYLPOPEVRYQALLKELIRNKARNQNCALLQEAQAVVVSALPQRAENKLHV 5880
Db 5821 PPPLOHLYLPOPEVRYQALLKELIRNKARNQNCALLQEAQAVVVSALPQRAENKLHV 5880
Qy 5881 SLMENYFGTLEALCEPIRQGHFIVWEGAPGARMWKGNHRVFLFRNHLVICKPRRDSRT 5940
Db 5881 SLMENYFGTLEALCEPIRQGHFIVWEGAPGARMWKGNHRVFLFRNHLVICKPRRDSRT 5940
Qy 5941 DTVSIVFRNMWKLSSIDLNDQVEGDDRAFEVWQREDSVRKYLLOARTAIKSSWVKIC 6000
Db 5941 DTVSIVFRNMWKLSSIDLNDQVEGDDRAFEVWQREDSVRKYLLOARTAIKSSWVKIC 6000
Qy 6001 GIQORLALPVWRPDPFEEELADCTABELGETVKLACRVGTGTPKPVISWYKDGKAVQVDPHH 6060
Db 6001 GIQORLALPVWRPDPFEEELADCTABELGETVKLACRVGTGTPKPVISWYKDGKAVQVDPHH 6060
Qy 6061 ILIEDPDGSCALILDSLTGVDSDGQYMCFAASAGNCSTLGKILVQVPPRFVNVKVRASPFV 6120
Db 6061 ILIEDPDGSCALILDSLTGVDSDGQYMCFAASAGNCSTLGKILVQVPPRFVNVKVRASPFV 6120

Qy	6121	EGEDAQFTCTIEGAPYQIWRKYKDGALLTGNKFKQTLSEBPSRGLLVLVIVIRAAASKEDLGLY	6180
Db	6121	EGEDAQFTCTIEGAPYQIWRKYKDGALLTGNKFKQTLSEBPSRGLLVIVIRAAASKEDLGLY	6180
Qy	6181	ECELVNRLGSARASAEURLIOSPMLQAOEQCHREQLVAAVDDTLERADQSVTSVLKRLLG	6240
Db	6181	ECELVNRLGSARASAEURLIOSPMLQAOEQCHREQLVAAVDDTLERADQSVTSVLKRLLG	6240
Qy	6241	PKAPGPGSTGDLTGPGPCPRGAPALQETGSPQPVVTGTSEAPAVPRVPQPLLHSGPQEPE	6300
Db	6241	PKAPGPGSTGDLTGPGPCPRGAPALQETGSPQPVVTGTSEAPAVPRVPQPLLHSGPQEPE	6300
Qy	6301	AIARAQEWTVPIRMEGAAWPGAGTGELLMDVHSHVHVRETTORTYTYQAIIDHTARPPSMQ	6360
Db	6301	AIARAQEWTVPIRMEGAAWPGAGTGELLMDVHSHVHVRETTORTYTYQAIIDHTARPPSMQ	6360
Qy	6361	VTIEDVQAOGTGTAQFAEITEGDPQPSVWYKQSVQLVDSITRUSQQOEGTTYSLVLRHVA	6420
Db	6361	VTIEDVQAOGTGTAQFAEITEGDPQPSVWYKQSVQLVDSITRUSQQOEGTTYSLVLRHVA	6420
Qy	6421	SKDAGVYVTCLAQNTGGQVLCKAELLVLGGDNEPDSEKQSHRRKLHSPYKEIGRGVFG	6480
Db	6421	SKDAGVYVTCLAQNTGGQVLCKAELLVLGGDNEPDSEKQSHRRKLHSPYKEIGRGVFG	6480
Qy	6481	FVKRVQHKGNKILCAAKFPIPLURSTRQAAYRERDILAAALSHPLVTGLLDQOFETRKTILILI	6540
Db	6481	FVKRVQHKGNKILCAAKFPIPLURSTRQAAYRERDILAAALSHPLVTGLLDQOFETRKTILILI	6540
Qy	6541	LELCSBELDLRLYRKGVWTEAEVKVYIQOLVEGLHYLHSHGVHLHLDIKPSNITLMVHPAR	6600
Db	6541	LELCSBELDLRLYRKGVWTEAEVKVYIQOLVEGLHYLHSHGVHLHLDIKPSNITLMVHPAR	6600
Qy	6601	EDIKICDFGAQNIITPAELOFSQXGSPFVSPELIIQONPVSEASDIWAMGVISYLSLTC	6660
Db	6601	EDIKICDFGAQNIITPAELOFSQXGSPFVSPELIIQONPVSEASDIWAMGVISYLSLTC	6660
Qy	6661	SPFAGESDRATLLNVLGRVSWSPMAAHLSEDAKDFIKATLQAPQAPRPSAAQCLSHPW	6720
Db	6661	SPFAGESDRATLLNVLGRVSWSPMAAHLSEDAKDFIKATLQAPQAPRPSAAQCLSHPW	6720
Qy	6721	FLKSWPAEEAHFINTKQLFLLABSRWORSIMSYKSLVMRSITPELLRGPDSPSLGVAR	6780
Db	6721	FLKSWPAEEAHFINTKQLFLLABSRWORSIMSYKSLVMRSITPELLRGPDSPSLGVAR	6780
Qy	6781	HLCRDTGCGSSSSSSSDNELAPFAFAKSLPPSPVTHSPHLLHPRGFLRPSASLPEEAEASE	6840
Db	6781	HLCRDTGCGSSSSSSSDNELAPFAFAKSLPPSPVTHSPHLLHPRGFLRPSASLPEEAEASE	6840
Qy	6841	RSTEAPAPPASPEGAGPAAQGCVPVPHSVIRSLFYHOAGSPBHGALAPGSRHRPARRRH	6900
Db	6841	RSTEAPAPPASPEGAGPAAQGCVPVPHSVIRSLFYHOAGSPBHGALAPGSRHRPARRRH	6900
Qy	6901	LLKGGYTAGALPGLREPLMEHRVLEEEAAREEQNTLLAKAPSPETALRLPASGTHLAPGH	6960
Db	6901	LLKGGYTAGALPGLREPLMEHRVLEEEAAREEQNTLLAKAPSPETALRLPASGTHLAPGH	6960
Qy	6961	SHSLEHSDPSTPRPSSBACGEAQLPAPGGGAPIRDMGHPPQGSKOLPSTGCHPGPTAQPE	7020
Db	6961	SHSLEHSDPSTPRPSSBACGEAQLPAPGGGAPIRDMGHPPQGSKOLPSTGCHPGPTAQPE	7020
Qy	7021	RPSPDSPMWQAPPFCHPKQGSAPQEGCSPHPAVAPCPPGSPFPPGSCKEAPLVPSPPFLGQ	7080
Db	7021	RPSPDSPMWQAPPFCHPKQGSAPQEGCSPHPAVAPCPPGSPFPPGSCKEAPLVPSPPFLGQ	7080
Qy	7081	POAPPAFAKASPPPLDSKMGPGDLSLQRPKPGPCSSPGSASQASSSQVSSRLRVGSSQVGT	7140
Db	7081	POAPPAFAKASPPPLDSKMGPGDLSLQRPKPGPCSSPGSASQASSSQVSSRLRVGSSQVGT	7140
Qy	7141	EPGPSLDAEGWTOEAEDLSNSTPTLQRPQCAOTWRKFSLGGRGYAGVACGYTFAFGGDA	7200
Db	7141	EPGPSLDAEGWTOEAEDLSNSTPTLQRPQCAOTWRKFSLGGRGYAGVACGYTFAFGGDA	7200
Qy	7201	GGMLGQGPMMWARIWAYSQSEEEEBQBARAESQSEEQEARAESPLPQVSARPVPEVGRA	7260

Db	7201	GGMLGQPMWARIATAWYQSSEEBEQEARAESQSEEQEARAPSPQVBARVPVEVGRA	7260
QY	7261	PTRSSPEPTWEDIGQVSLVQIRDLSGDAAADTISLIDISEVDPAYLNLSDLYDKYLPF	7320
Db	7261	PTRSSPEPTWEDIGQVSLVQIRDLSGDAAADTISLIDISEVDPAYLNLSDLYDKYLPF	7320
QY	7321	EFMIFRKVPKSAQPEPPSPMAEBELAEFPPTWPGELGPHAGLEITBESSEDVALLAE	7380
Db	7321	EFMIFRKVPKSAQPEPPSPMAEBELAEFPPTWPGELGPHAGLEITBESSEDVALLAE	7380
QY	7381	AAVGRKKWSPSRSLFHPGRHLPLDEPAELGLRERVKASVSHISRLIKGRPEGLEKEG	7440
Db	7381	AAVGRKKWSPSRSLFHPGRHLPLDEPAELGLRERVKASVSHISRLIKGRPEGLEKEG	7440
QY	7441	PPRKPKGLAGFRLSGLKSWDRATFFRELSDETWVLGQSVTLACQVSAQPAQAATWSKOG	7500
Db	7441	PPRKPKGLAGFRLSGLKSWDRATFFRELSDETWVLGQSVTLACQVSAQPAQAATWSKOG	7500
QY	7501	APLESSSRVLISATLKNFQLLTLLVVAEDLVYTCVSNALGTVTVTGVLKRAERPSSS	7560
Db	7501	APLESSSRVLISATLKNFQLLTLLVVAEDLVYTCVSNALGTVTVTGVLKRAERPSSS	7560
QY	7561	PCPDIGEVYADGVLLVWKPVESYGPVTYIYVQCSLEGGSWTTLASDIFDCCYLTSKLSRGG	7620
Db	7561	PCPDIGEVYADGVLLVWKPVESYGPVTYIYVQCSLEGGSWTTLASDIFDCCYLTSKLSRGG	7620
QY	7621	TYTPRTACVSKAGMGPPSYSPSEOVLLGCGPSHLASEEBSQGRSAQPLPSTKTTFAFQTOIQR	7680
Db	7621	TYTPRTACVSKAGMGPPSYSPSEOVLLGCGPSHLASEEBSQGRSAQPLPSTKTTFAFQTOIQR	7680
QY	7681	GRFSVVRQCKEASGRALAAKIIIPYHPKOKTAVLREYEAALKGLRHPHPLAQLHAAYLSPRH	7740
Db	7681	GRFSVVRQCKEASGRALAAKIIIPYHPKOKTAVLREYEAALKGLRHPHPLAQLHAAYLSPRH	7740
QY	7741	LVLILELCSGPPELLPCLAEASYSSEVKDYLNQMLSATQYLNQHLHLDLRSENMIIT	7800
Db	7741	LVLILELCSGPPELLPCLAEASYSSEVKDYLNQMLSATQYLNQHLHLDLRSENMIIT	7800
QY	7801	EYNLLKVVDLGNQSLQSEKVLPSDKPKDYLETWAPELLEGGQAVPOTDIIWALGVTAFIG	7860
Db	7801	EYNLLKVVDLGNQSLQSEKVLPSDKPKDYLETWAPELLEGGQAVPOTDIIWALGVTAFIG	7860
QY	7861	LSAEYPVSSSEGARDLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQPMGRPCASSCLQ	7920
Db	7861	LSAEYPVSSSEGARDLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQPMGRPCASSCLQ	7920
QY	7921	CPWLTEGPGACSRPAPVTFPTARLVRVVRNREKRALLYKRHNLAQVR	7968
Db	7921	CPWLTEGPGACSRPAPVTFPTARLVRVVRNREKRALLYKRHNLAQVR	7968
RESULT 2			
ADJ70485			
ID	ADJ70485	standard; protein; 6620 AA.	
XX			
AC	ADJ70485;		
XX			
DT	06-MAY-2004	(first entry)	
XX			
DE		Human heat mitochondrial protein as a therapeutic target SeqID2291.	
XX			
KW		mitochondrial; human; screening assay; diabetes mellitus;	
KW		Huntington's disease; osteoarthritis;	
KW		Leber's hereditary optic neuropathy; LHON;	
KW		mitochondrial encephalopathy lactic acidosis and stroke; MELAS;	
KW		myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;	
KW		neuroprotective; nontopic; antidiabetic;	
KW		osteopathic; ophthalmological; cytostatic.	
OS		Homo sapiens.	
XX			
FN	WO2003087768-A2.		

XX 23-OCT-2003.
XX 04-APR-2003; 2003WO-US010870.
XX 12-APR-2002; 2002US-0372843P.
PR 17-JUN-2002; 2002US-0389987P.
PR 20-SEP-2002; 2002US-0412418P.
XX (MITO-) MITOKOR.
PA (BUCK-) BUCK INST AGE RES.
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI Warnock DB;
XX WPI; 2003-845369/78.
XX Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
XX Claim 1; SEQ ID NO 2291; 180pp; English.
XX This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC cephalopathy, lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, neurotropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmologic and
CC cytoskeletal activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX Sequence 6620 AA;
Query Match 77.9%; Score 32134; DB 7; Length 6620;
Best Local Similarity 98.1%; Pred. NO. 0;
Matches 6250; Conservative 15; Mismatches 57; Indels 48; Gaps 9;
Qy 1 MDQPFSGAPRFLTRPKAFVSVGKDATLSQIVGNPTPOVSWKDDQOPVTAGARFLAQ 60
Db 1 MDQPFSGAPRFLTRPKAFVSVGKDATLSQIVGNPTPOVSWKDDQOPVTAGARFLAQ 60
Qy 61 DGDLYRLTILDLALGDSQYVCRARNATGEAFVAVGLQVDAEAAEQAHPFLRLPTSIR 120
Db 61 DGDLYRLTILDLALGDSQYVCRARNALGEAFVAVGLQVDAEAAEQAHPFLRLPTSIR 120
Qy 121 VREGSEATFRCKVGSPPRPAVSWKDGRRLEGPDPGRVVRVEELGEASALRIRAAARPDGG 180
Db 121 VREGSEATFRCKVGSPPRPAVSWKDGRRLEGPDPGRVVRVEELGEASALRIRAAARPDGG 180
Qy 181 TYEVRANPLGAASAAALVDSDAADTASBPSTSTAALLAHLQRRREAMEAGAPASP 240
Db 181 TYEVRANPLGAASAAALVDSDAADTASBPSTSTAALLAHLQRRREAMEAGAPASP 240
Qy 241 STGTRTCTVTGKHARLSQYVTEGPKPTVMKKDQQLVTEGRRHVVEDAQENFVLKILF 300
Db 241 STGTRTCTVTGKHARLSQYVTEGPKPTVMKKDQQLVTEGRRHVVEDAQENFVLKILF 300
Qy 301 CKQSDRGLYTCTASNLVQGTYSVLVVRREPAVFPKQLDLVREKESATFLCEVPQPS 360
Db 301 CKQSDRGLYTCTASNLVQGTYSVLVVRREPAVFPKQLDLVREKESATFLCEVPQPS 360
Qy 361 TEAAWFKETRLWASAKYIEEGTERLTVRNVSADDDAVYICETPEGSRTVAELAVQG 420
Db 361 TEAAWFKETRLWASAKYIEEGTERLTVRNVSADDDAVYICETPEGSRTVAELAVQG 420

Qy 421 NLLRKLPRKTAVRVGDVTAMFCVELAVPVPVHVLNRNQEEVVAGGRVAISAEGRHTLTIS 480
Db 421 NLLRKLPRKTAVRVGDVTAMFCVELAVPVPVHVLNRNQEEVVAGGRVAISAEGRHTLTIS 480
Qy 481 QCCLEDVGQVAFMAGDCQSTSTRFCVSAPRRPPLQPPVDVVPVKARMESSVILSWPPPHGE 540
Db 481 QCCLEDVGQVAFMAGDCQSTSTRFCVSAPRRPPLQPPVDVVPVKARMESSVILSWPPPHGE 540
Qy 541 RPTVIDGVLVEKKLGYTTWIRCHEAEWATPELTVDVABEGNFQFRVVSALNFGQSPY 600
Db 541 RPTVIDGVLVEKKLGYTTWIRCHEAEWATPELTVDVABEGNFQFRVVSALNFGQSPY 600
Qy 601 LEFPGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDLTVASAGEWFLDQALKASSVYBIH 660
Db 601 LEFPGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDLTVASAGEWFLDQALKASSVYBIH 660
Qy 661 CDRTRHTLTIREVPASLHGALQKPVANGIESSIRMEVRAAPCLTANKPAAAAAEVLARL 720
Db 661 CDRTRHTLTIREVPASLHGALQKPVANGIESSIRMEVRAAPCLTANKPAAAAAEVLARL 720
Qy 721 HEAQLLAELSDDAAAATWLDKGRITLSPGPKYEQASAGRRVLLVRDVARDDAGLYECVS 780
Db 721 HEAQLLAELSDDAAAATWLDKGRITLSPGPKYEQASAGRRVLLVRDVARDDAGLYECVS 780
Qy 781 RGGRIAYQLSVQGLARFLHKDMAGSCYDAVAGGPAQFECETSEAHVHVHVKDGMELGHS 840
Db 781 RGGRIAYQLSVQGLARFLHKDMAGSCYDAVAGGPAQFECETSEAHVHVHVKDGMELGHS 840
Qy 841 GERFLOEDVGTTRHLVAATVTRQDEGTYSCVGDSDVDRLRVSEPKVVFVFAKEQLARLKL 900
Db 841 GERFLOEDVGTTRHLVAATVTRQDEGTYSCVGDSDVDRLRVSEPKVVFVFAKEQLARLKL 900
Qy 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVCMEATGCTRRLLVVOQAGADAGEY 960
Db 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVCMEATGCTRRLLVVOQAGADAGEY 960
Qy 961 SCEAGGQRLSFHLDVKEPKVVFVFAKDQVAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
Db 961 SCEAGGQRLSFHLDVKEPKVVFVFAKDQVAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
Qy 1021 KLSSSLKHVVEAKCRRRLVVOQAGKTDAGDYSCEARGQVSFRLHITPEKMFVAKESQV 1080
Db 1021 KLSSSLKHVVEAKCRRRLVVOQAGKTDAGDYSCEARGQVSFRLHITPEKMFVAKESQV 1080
Qy 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSKVCMEATGCTRRLLVVOQAGKAD 1140
Db 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSKVCMEATGCTRRLLVVOQAGKAD 1140
Qy 1141 AGEYSCBAGGQVSFHLHITPEKGVFAKEQSVHNEVQAEAGTATMLSCVAQOPQTEVTWY 1200
Db 1141 AGEYSCBAGGQVSFHLHITPEKGVFAKEQSVHNEVQAEAGTATMLSCVAQOPQTEVTWY 1200
Qy 1201 KDGKLSKSSSKVRMEVKGCTRRLLVVOQKADAGEYSCBAGQVSFOLHITPEKAVFAK 1260
Db 1201 KDGKLSKSSSKVRMEVKGCTRRLLVVOQKADAGEYSCBAGQVSFOLHITPEKAVFAK 1260
Qy 1261 EQLVHNEVTRTEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAGCWRQLVVOQA 1320
Db 1261 EQLVHNEVTRTEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAGCWRQLVVOQA 1320
Qy 1321 GQADAGAYTCBAGQRLSFHLDVSEPKAVFAKEQLAHLKVAEAGAIATLSCEVAQAQTE 1380
Db 1321 GQADAGAYTCBAGQRLSFHLDVSEPKAVFAKEQLAHLKVAEAGAIATLSCEVAQAQTE 1380
Qy 1381 VTMVYKDGKLSKSSSKVRMEVKGCTRRLLVVOQACQADTGEYSCBAGGQRLSFSLDVAEPKV 1440
Db 1381 VTMVYKDGKLSKSSSKVRMEVKGCTRRLLVVOQACQADTGEYSCBAGGQRLSFSLDVAEPKV 1440
Qy 1441 VFAKEQPVHREVOQAAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRMEVKGCTRRLLV 1500
Db 1441 VFAKEQPVHREVOQAAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRMEVKGCTRRLLV 1500

QY 1501 VQAGQADAGEYCEAGSORLSFHLHVAEPKAVPAKEQASREVQAEACTSATLSCEVAQ 1560
DB 1501 VQAGQADAGEYCEAGSORLSFHLHVAEPKAVPAKEQASREVQAEACTSATLSCEVAQ 1560
QY 1561 AQTEVTWYKDGKLLSSSKVRMEAVGCTRRLLVYQAGQADAGEYCEAGQDQRLSFLHVA 1620
DB 1561 AQTEVTWYKDGKLLSSSKVRMEAVGCTRRLLVYQAGQADAGEYCEAGQDQRLSFLHVA 1620
QY 1621 EPKVVPKAEQPAHREVQAEAGSATLSCEVAQAQTEVTWYKDGKLLSSSKVRVBAVCGT 1680
DB 1621 EPKVVPKAEQPAHREVQAEAGSATLSCEVAQAQTEVTWYKDGKLLSSSKVRVBAVCGT 1680
QY 1681 RRLVYQAGQADAGEYCEAGQRLSFRHLHVAELEPOJSERPCREPLVVKHEHDIILTA 1740
DB 1681 RRLVYQAGQADAGEYCEAGQRLSFRHLHVAELEPOJSERPCREPLVVKHEHDIILTA 1740
QY 1741 TLATPSAATVTLWKDGVIRRSKRHETASQGDTHLTVHGAQVLDLSAIYSCRVAEGQDF 1800
DB 1741 TLATPSAATVTLWKDGVIRRSKRHETASQGDTHLTVHGAQVLDLSAIYSCRVAEGQDF 1800
QY 1801 PVQVEEVAACFRLLEPVYCGELGGTUTLACELSPACAEVWRCGNTQPRVGRFQWVAEG 1860
DB 1801 PVQVEEVAACFRLLEPVYCGELGGTUTLACELSPACAEVWRCGNTQPRVGRFQWVAEG 1860
QY 1861 PVRSLTVLGLRAEDAGEYVCESRDDHTSAQLTVSVPRVVKFMSGLSVVABEGGEATFQC 1920
DB 1861 PVRSLTVLGLRAEDAGEYVCESRDDHTSAQLTVSVPRVVKFMSGLSVVABEGGEATFQC 1920
QY 1921 VYPSDVAVWFRDQALLQPEKFAISQSGASHSLTISDLVLEDAGQITVBAEGASSAA 1980
DB 1921 VYPSDVAVWFRDQALLQPEKFAISQSGASHSLTISDLVLEDAGQITVBAEGASSAA 1980
QY 1981 LVRAPVLFKKLBPQTVBERSSVTLEVELTRPWPPELRWTRNATALAPGKNVEIHAEGA 2040
DB 1981 LVRAPVLFKKLBPQTVBERSSVTLEVELTRPWPPELRWTRNATALAPGKNVEIHAEGA 2040
QY 2041 RHRLVLHNVGFADRGFFCETPDDTKOAKLTVMQVRLVRGLQAVEAREOGTATMEVOL 2100
DB 2041 RHRLVLHNVGFADRGFFCETPDDTKOAKLTVMQVRLVRGLQAVEAREOGTATMEVOL 2100
QY 2101 SHADVDGSWTRDGLRFQOQPTCHLAVRGPMTHTLTLGLRPEDSGLMVFKAEVHTSARLV 2160
DB 2101 SHADVDGSWTRDGLRFQOQPTCHLAVRGPMTHTLTLGLRPEDSGLMVFKAEVHTSARLV 2160
QY 2161 VTELVPVSRPLQDVVTTKEKVTLCELSRPNVDVRLKDGVELRAGKTMAIAAQGACR 2220
DB 2161 VTELVPVSRPLQDVVTTKEKVTLCELSRPNVDVRLKDGVELRAGKTMAIAAQGACR 2220
QY 2221 SLTIYRCBFADQGVVCDADHAQSSASVKVQGRVTLIYRRLAEDAGEIQFVAENAESR 2280
DB 2221 SLTIYRCBFADQGVVCDADHAQSSASVKVQGRVTLIYRRLAEDAGEIQFVAENAESR 2280
QY 2281 AOLRVKELPVTILVRPLDKIAEKHGRVLECOVSRASQVRWFKGSQELQPGPKYELVSD 2340
DB 2281 AOLRVKELPVTILVRPLDKIAEKHGRVLECOVSRASQVRWFKGSQELQPGPKYELVSD 2340
QY 2341 GLYRKLIIISDVHAEDEDTYTCADGVKTSAGQFVEEQSITIVRGLODVTVMPEPAPWPEC 2400
DB 2341 GLYRKLIIISDVHAEDEDTYTCADGVKTSAGQFVEEQSITIVRGLODVTVMPEPAPWPEC 2400
QY 2401 ETSIPSVRPPKWLKGKTVLQAGNVVLEQECTVHRLMLRRTCTMTGPHFTVVGSRSSA 2460
DB 2401 ETSIPSVRPPKWLKGKTVLQAGNVVLEQECTVHRLMLRRTCTMTGPHFTVVGSRSSA 2460
QY 2461 RLWVSDIPVILTRPLEPTKRELOQSVLSCDPRPAPKAVQWYKDDTPLSPSEKFKWLSLEG 2520
DB 2461 RLWVSDIPVILTRPLEPTKRELOQSVLSCDPRPAPKAVQWYKDDTPLSPSEKFKWLSLEG 2520
QY 2521 QMAELRIURLMPADAGVYRCQAGSAHSTEVTVAREVTVTGPLQDAEATEEGWASPSCE 2580
DB 2521 QMAELRIURLMPADAGVYRCQAGSAHSTEVTVAREVTVTGPLQDAEATEEGWASPSCE 2580
QY 2581 LSHDEEVEVWSLNGHPLVNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLUKVTSAKL 2640

DB 2581 LSHDEEVEVWSLNGHPLVNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLUKVTSAKL 2640
QY 2641 EVRVKPVVFLKALDDLSAEBRGTTALQCEVSDPEAHVWVRKDGVLQGPSDKYDFLHTAGT 2700
DB 2641 EVRVKPVVFLKALDDLSAEBRGTTALQCEVSDPEAHVWVRKDGVLQGPSDKYDFLHTAGT 2700
QY 2701 RGLVVDVSPEDAGLYTCHVGSSETRARVRVHDLHVGITKRLKTMVEVLEGBSCSCECVLS 2760
DB 2701 RGLVVDVSPEDAGLYTCHVGSSETRARVRVHDLHVGITKRLKTMVEVLEGBSCSCECVLS 2760
QY 2761 HESASDPAMVTVGGKTGSSSRFOATROGRKYILVVRREAAAPSDAGEVVSVRGLTSKASL 2820
DB 2761 HESASDPAMVTVGGKTGSSSRFOATROGRKYILVVRREAAAPSDAGEVVSVRGLTSKASL 2820
QY 2821 IVRERPAAIIPKLEDDQWVAPGEDVELRCELRSRAGTPVHMLKDKRAIKRSQKYDVVCEGTM 2880
DB 2821 IVRERPAAIIPKLEDDQWVAPGEDVELRCELRSRAGTPVHMLKDKRAIKRSQKYDVVCEGTM 2880
QY 2881 AMLVIRGASLKDAGEYTCVEASASTASLHVEKANCFTBELTNLOVEEKGTAFTVCTKTE 2940
DB 2881 AMLVIRGASLKDAGEYTCVEASASTASLHVEKANCFTBELTNLOVEEKGTAFTVCTKTE 2940
QY 2941 HPAATVTVWRKGLLELRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGQASQAQLLV 3000
DB 2941 HPAATVTVWRKGLLELRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGQASQAQLLV 3000
QY 3001 QGRRVHIIEDLEDVDVQEGSSATFRCRISPANYEPVHWFLLDKPTLHANELNEIDAQPGGY 3060
DB 3001 QGRRVHIIEDLEDVDVQEGSSATFRCRISPANYEPVHWFLLDKPTLHANELNEIDAQPGGY 3060
QY 3061 HVLTLRQALXKDSGTIYFEAGDQASAAALRYTERKPSVFSRELTDTATITEGEDTLVLCETS 3120
DB 3061 HVLTLRQALXKDSGTIYFEAGDQASAAALRYTERKPSVFSRELTDTATITEGEDTLVLCETS 3120
QY 3121 TCDIPMCWTKDGKTLRGSARCOLSHEGHRALQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180
DB 3121 TCDIPMCWTKDGKTLRGSARCOLSHEGHRALQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180
QY 3181 HARVPRFOBALKDLEVLGGAAATLRCVLSVAAAPVKWCYGNVLRPDKYSLRQSGAMLE 3240
DB 3181 HARVPRFOBALKDLEVLGGAAATLRCVLSVAAAPVKWCYGNVLRPDKYSLRQSGAMLE 3240
QY 3241 LVVNRRLRQDSGRYSCSFGDQTTSATLTVTALPAQFIGLKNKEATEGATATLRCELSKT 3300
DB 3241 LVVNRRLRQDSGRYSCSFGDQTTSATLTVTALPAQFIGLKNKEATEGATATLRCELSKT 3300
QY 3301 APVEMWKGSETLRDGDYCLRDQGMCELQIRGLAMVDAAEYSCVCGBERTSASLTIIRPM 3360
DB 3301 APVEMWKGSETLRDGDYCLRDQGMCELQIRGLAMVDAAEYSCVCGBERTSASLTIIRPM 3360
QY 3361 PAHFIGRLRHQESSIEGATATLRCELSKAAPVEMWKGRESLRDGRHSLRQDGAVALCELOQC 3420
DB 3361 PAHFIGRLRHQESSIEGATATLRCELSKAAPVEMWKGRESLRDGRHSLRQDGAVALCELOQC 3420
QY 3421 GLAVADAGEYSCVCGBERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWELSKVAPVE 3480
DB 3421 GLAVADAGEYSCVCGBERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWELSKVAPVE 3480
QY 3481 WRKGPENLRDGDYTLRQEGTRCELOICGLAMADAGEYLCVCGBERTSATLTIRALPARF 3540
DB 3481 WRKGPENLRDGDYTLRQEGTRCELOICGLAMADAGEYLCVCGBERTSATLTIRALPARF 3540
QY 3541 IEDVKNQREAREGATAVLQCELSNAAPVEMWKGSETLRDGRYSLRQDGTKCELOIRGLAM 3600
DB 3541 IEDVKNQREAREGATAVLQCELSNAAPVEMWKGSETLRDGRYSLRQDGTKCELOIRGLAM 3600
QY 3601 ADTGEYSVCVCGBERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCBLSKMAPVEMWKG 3660
DB 3601 ADTGEYSVCVCGBERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCBLSKMAPVEMWKG 3660
QY 3661 HETLRDGRHSLRDQGARCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPSPFIEGL 3720

Db 3661 HETLRDGRHSURQDARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSKFI EBL 3720
Qy 3721 RNEEATEGDTATLWCELSKAAAPVEMRKGHETLRDGRHSRLRQDGRCELOIRGLAVVDAG 3780
Db 3721 RNEEATEGDTATLWCELSKAAAPVEMRKGHETLRDGRHSRLRQDGRCELOIRGLAVVDAG 3780
Qy 3781 EYSCVCGQERTSATLTVRALPARFI EDVKNOBAREGATAVLQCELSKAAPVEMRKGS ETL 3840
Db 3781 EYSCVCGQERTSATLTVRALPARFI EDVKNOBAREGATAVLQCELSKAAPVEMRKGS ETL 3840
Qy 3841 RCGDRYSRLRQDTRCELOIHGLSVADTGEYSVCVCGQERTSATLTVRALPOVPREPLQSLQ 3900
Db 3841 RCGDRYSRLRQDTRCELOIHGLSVADTGEYSVCVCGQERTSATLTVRALPOVPREPLQSLQ 3900
Qy 3901 ABEGSTATLQCELSBPTATVWWSKGLQLOANGREPRLOQCTAELVLDIQLQREDTG EYT 3960
Db 3901 ABEGSTATLQCELSBPTATVWWSKGLQLOANGREPRLOQCTAELVLDIQLQREDTG EYT 3960
Qy 3961 CTCGSOATSATLTVTAAPVRFURELQHQEVDEGGTAHLCCELSRAGASVEMRKGSLOLFP 4020
Db 3961 CTCGSOATSATLTVTAAPVRFURELQHQEVDEGGTAHLCCELSRAGASVEMRKGSLOLFP 4020
Qy 4021 CAKYQMWQDGAABLLVRGVEQEDAGDYTCDTGHTQSMASLSVRPRPKFKTRLOSLEQ E 4080
Db 4021 CAKYQMWQDGAABLLVRGVEQEDAGDYTCDTGHTQSMASLSVRPRPKFKTRLOSLEQ E 4080
Qy 4081 TGDIALRCCQLSDAESGAVVOWLKEGVELHAGPKYEMRSQGA TRELLIHOLEAKDTGEYA 4140
Db 4081 TGDIALRCCQLSDAESGAVVOWLKEGVELHAGPKYEMRSQGA TRELLIHOLEAKDTGEYA 4140
Qy 4141 CVTGGQKTAASLRVTEPEVTI VRGLVDAEVTADEDEVFSCVSRAGATGVQWCLQGLP LQ 4200
Db 4141 CVTGGQKTAASLRVTEPEVTI VRGLVDAEVTADEDEVFSCVSRAGATGVQWCLQGLP LQ 4200
Qy 4201 SNEVTEVAVRGR IHTLRLKGVTPEDAGTVSPHLGNHASSAQLTVRAPEVTILEPLQDVQ 4260
Db 4201 SNEVTEVAVRGR IHTLRLKGVTPEDAGTVSPHLGNHASSAQLTVRAPEVTILEPLQDVQ 4260
Qy 4261 LSEGDQASFQCLSRASQEARWALGGVPLQANEMNDITVFGTLLHLTLHKVTLEDA GT 4320
Db 4261 LSEGDQASFQCLSRASQEARWALGGVPLQANEMNDITVFGTLLHLTLHKVTLEDA GT 4320
Qy 4321 VSFHVGTCSBAQLKVTAKNTVVRGLENVLEALEGGEALFECQLSOPVEAAHTWLLDDEP V 4380
Db 4321 VSFHVGTCSBAQLKVTAKNTVVRGLENVLEALEGGEALFECQLSOPVEAAHTWLLDDEP V 4380
Qy 4381 RTSNAEVVFRENGIRHLLLLKNLRPQDS CRVTFLAGDMVTSAFITVRGWRLEILEPLKN 4440
Db 4381 RTSNAEVVFRENGIRHLLLLKNLRPQDS CRVTFLAGDMVTSAFITVRGWRLEILEPLKN 4440
Qy 4441 AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQPDSDWTADGSHQALLLSAQOPHH 4500
Db 4441 AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQPDSDWTADGSHQALLLSAQOPHH 4500
Qy 4501 AGEVTFACRDVAVASARLTVLGLPDPEDA EVVAHSSHTVTLSWAAPMSDGGGCGYRVE 4560
Db 4501 AGEVTFACRDVAVASARLTVLGLPDPEDA EVVAHSSHTVTLSWAAPMSDGGGCGYRVE 4560
Qy 4561 VKEGATGQWRLCHELVPGPECVVDGLAPGETYRFRVAAGVPVGA GEPVHLPTVRLAEP P 4620
Db 4561 VKEGATGQWRLCHELVPGPECVVDGLAPGETYRFRVAAGVPVGA GEPVHLPTVRLAEP P 4620
Qy 4621 KPVPPQSPAPSRQVAAGEDVLSLEVVAEAGEVIWHKGMERIO PGGRFEVVSQGRQOML 4680
Db 4621 KPVPPQSPAPSRQVAAGEDVLSLEVVAEAGEVIWHKGMERIO PGGRFEVVSQGRQOML 4680
Qy 4681 VIKGFTAEDQGEYHCGGLAQGSI CPAAATFQVALSPASVDEAPQPSLPPEAAQEGDLHL L 4740
Db 4681 VIKGFTAEDQGEYHCGGLAQGSI CPAAATFQVALSPASVDEAPQPSLPPEAAQEGDLHL L 4740
Qy 4741 EALARKRMSREPTLDSISELPEEDGRSQRLPQABEEVAPDLS EGYSTADELARTGDADL 4800
Db 4741 EALARKRMSREPTLDSISELPEEDGRSQRLPQABEEVAPDLS EGYSTADELARTGDADL 4800

Qy 4801 SHTSDDESRACTPSLVTYLLKKAGRPGTSP LASKVGAPAAPSVKPOQOQEPLAAVRPPLG 4860
Db 4801 SHTSDDESRACTPSLVTYLLKKAGRPGTSP LASKVGAPAAPSVKPOQOQEPLAAVRPPLG 4860
Qy 4861 DLSTKDLGDPMSDKAAAVKIQAAAFKGYKVRKEMKQOEGPMFSHTFGDTBAQVGDALRLSCV 4920
Db 4861 DLSTKDLGDPMSDKAAAVKIQAAAFKGYKVRKEMKQOEGPMFSHTFGDTBAQVGDALRLSCV 4920
Qy 4921 VASKADYRARWLKDGVELTDGRHHHIDOLG DGTCSLLIAGLDRADAGCYTCOVSNKFGQV 4980
Db 4921 VASKADYRARWLKDGVELTDGRHHHIDOLG DGTCSLLIAGLDRADAGCYTCOVSNKFGQV 4980
Qy 4981 THSACVVVSGSESAESSGGELDDAFRAARRLHRLFR TKSPA EVSDEELFLSADEGPA 5040
Db 4981 THSACVVVSGSESAESSGGELDDAFRAARRLHRLFR TKSPA EVSDEELFLSADEGPA 5040
Qy 5041 EPPEPADQWTVREDEHFCIRFEALTEARQAVTRFQEMFATLGI GVEIKLVEQGPRR VEM 5100
Db 5041 EPPEPADQWTVREDEHFCIRFEALTEARQAVTRFQEMFATLGI GVEIKLVEQGPRR VEM 5100
Qy 5101 CISKETPAPVVPPEPLSLTSDAAPVFLTELQNEQVODGYPVSPDCVVTGQMPSPVRWF 5160
Db 5101 CISKETPAPVVPPEPLSLTSDAAPVFLTELQNEQVODGYPVSPDCVVTGQMPSPVRWF 5160
Qy 5161 KDGKLLBEDDHVMINEQQQGHQLIITAVVPADMGVVRC LAENSMGVSSTKAE LRVDLTS 5220
Db 5161 KDGKLLBEDDHVMINEQQQGHQLIITAVVPADMGVVRC LAENSMGVSSTKAE LRVDLTS 5220
Qy 5221 TDYTAADATESSYFSAQGYLSSREOEGTSTTDEGOLPOVVBELRDLQVAPGTR LAKF 5280
Db 5221 TDYTAADATESSYFSAQGYLSSREOEGTSTTDEGOLPOVVBELRDLQVAPGTR LAKF 5280
Qy 5281 QLKVKGYPA RLYWFKDGOPLTASAHIRMTDKKILHTLEIISVTR EDSGOYAAVISNMG 5340
Db 5281 QLKVKGYPA RLYWFKDGOPLTASAHIRMTDKKILHTLEIISVTR EDSGOYAAVISNMG 5340
Qy 5341 AAYSSARLLVRGPDEPEKEPASDVHEQLVPPRMLERFTPKKKVKGSSITTSVKVEGR PVP 5400
Db 5341 AAYSSARLLVRGPDEPEKEPASDVHEQLVPPRMLERFTPKKKVKGSSITTSVKVEGR PVP 5400
Qy 5401 TVHMLREABERGVLWIGPDTPGYTVASSAQS HSLVLLDVGRHQGTTCIASNAAGQALC 5460
Db 5401 TVHMLREABERGVLWIGPDTPGYTVASSAQS HSLVLLDVGRHQGTTCIASNAAGQALC 5460
Qy 5461 SASLHVSGLPKVERQEKKEALISTFLOGTTQATISAQGLETASPADLGGOKKEP LAAKE 5520
Db 5461 SASLHVSGLPKVERQEKKEALISTFLOGTTQATISAQGLETASPADLGGOKKEP LAAKE 5520
Qy 5521 ALGHLSLAEVGTBEFLQKLT SQITEMVSAKITOAKLQVP GGDSDSDSKTSPASPRHGRSR 5580
Db 5521 ALGHLSLAEVGTBEFLQKLT SQITEMVSAKITOAKLQVP GGDSDSDSKTSPASPRHGRSR 5580
Qy 5581 PSSSIQSSSESDGDARGEI FDIYVVTADYLPUGASQDAILTREGQVVEVLDAAHPLRW 5640
Db 5581 PSSSIQSSSESDGDARGEI FDIYVVTADYLPUGASQDAILTREGQVVEVLDAAHPLRW 5640
Qy 5641 LVRTKPTKSSPSRGQWSPAYLDRRLKLSPEWGAABEPPEPGEAVSDEYKARLSSVIOE 5700
Db 5641 LVRTKPTKSSPSRGQWSPAYLDRRLKLSPEWGAABEPPEPGEAVSDEYKARLSSVIOE 5700
Qy 5701 LLSSEQAFVBELOFLQSHHLLQHLERCPHPVPIA VAGQKAVIPRNVRD IGRPHSSFLQELQQ 5760
Db 5701 LLSSEQAFVBELOFLQSHHLLQHLERCPHPVPIA VAGQKAVIPRNVRD IGRPHSSFLQELQQ 5760
Qy 5761 CDTDDDDVAMCFIKNOAAFEQYLEFLVGRVQAESV VVWSTAIQEFYKKYKAEALLAGDPSQP 5820
Db 5761 CDTDDDDVAMCFIKNOAAFEQYLEFLVGRVQAESV VVWSTAIQEFYKKYKAEALLAGDPSQP 5820
Qy 5821 PPPPLQHYLSQPVVERVORYQALLKELRNKARNQNCALL EOA VAVVSALLPQRAENKLHV 5880
Db 5821 PPPPLQHYLSQPVVERVORYQALLKELRNKARNQNCALL EOA VAVVSALLPQRAENKLHV 5880

QY 5881 SLWENY PGTLEALGPPIQGHFIVMEGAPCBMPKGNHNVFLFRNHLVICKPRDSRT 5940
 Db 5881 SLWENY PGTLEALGPPIQGHFIVMEGAPCBMPKGNHNVFLFRNHLVICKPRDSRT 5940
 QY 5941 DTVSVFRNMKLSIDLDQVEGDDRAFEVWQEREDSVRKYLLOARTAIKSSWVKEIC 6000
 Db 5941 DTVSVFRNMKLSIDLDQVEGDDRAFEVWQEREDSVRKYLLOARTAIKSSWVKEIC 6000
 QY 6001 GIQORLALPVWRPPPEBELADCTAELGETVTKLACRVGTGTPKPVISWYKDGKAVQVDPHH 6060
 Db 6001 GIQORLALPVWRPPPEBELADCTAELGETVTKLACRVGTGTPKPVISWYKDGKAVQVDPHH 6060
 QY 6061 ILIEDPDGSCALILDSLTGVDSSQYMCFAASAGNCSTLGLKTLVQVPRFVNKVRASPV 6120
 Db 6061 ILIEDPDGSCALILDSLTGVDSSQYMCFAASAGNCSTLGLKTLVQVPRFVNKVRASPV 6120
 QY 6121 EGEDAQFTCTIEGAPYQIRWYKDGALLTGNKFTLSEPRSGLLVLVIRAAKEDLGLY 6180
 Db 6121 EGEDAQFTCTIEGAPYQIRWYKDGALLTGNKFTLSEPRSGLLVLVIRAAKEDLGLY 6180
 QY 6181 ECELVNRLGASARASAEIRIQSPMLQAQEQCHREQLVAIVEDTTLER-----ADQEV 6232
 Db 6181 ECELVNRLGASARASAEIRIQSPMLQAQEQCHREQLVAIVEDTTLER-----ADQEV 6232
 QY 6233 SVLKRLGAPKAPGSTGDLTGPGCPRG-----APAL-----QETGSP- 6271
 Db 6241 TVVKSPPQRR--SPKSPSPSPSCASPLRPLGLLADPLLYPGAGQRRPRAEPGQKPV 6299
 QY 6272 -PVTGTSAPAVPPRV-----PQPLLHGPE-----QEPEAIARAQEWTPIRMGA 6317
 Db 6300 VPTLVVTEAEAHSPALPGLSGPQPKWVEETIEVRVKMGQGVSPTE--VPRSSGH 6357
 QY 6318 AW--PGAGTG 6325
 Db 6358 LFTLPGATPG 6367

RESULT 3

ABG76186
 ID ABG76186 standard; protein; 2630 AA.
 XX
 AC ABG76186;
 DT 09-MAY-2003 (first entry)
 XX
 DE Human serine/threonine or protein kinase 59079.
 KW Human; enzyme; serine/threonine kinase; protein kinase; 59079;
 KW cardiovascular disease; heart failure; myocardial infarction;
 KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma; immunogen;
 KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
 KW haemolytic anaemia; cellular proliferative disorder; cancer;
 KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
 KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
 KW multiple sclerosis.
 XX
 OS Homo sapiens.
 XX
 PN US2002168742-A1.
 XX
 PD 14-NOV-2002.
 XX
 PF 15-FEB-2002; 2002US-00077130.
 XX
 PR 15-FEB-2001; 2001US-0269201P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Kapeller-Libermann R, Acton SL;
 XX
 DR WPI; 2003-298729/29.
 DR N-PSDB; ABX11641.
 XX

PT Novel isolated human protein kinase, designated 59079 or 12599
 PT polypeptide, useful as diagnostic and therapeutic agents for preventing
 PT cardiovascular diseases, proliferative disorders, and protein kinase
 PT disorders.
 XX
 PS Claim 8; Page 48-54; 119pp; English.
 XX
 CC The invention relates to an isolated human serine/threonine or protein
 CC kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
 CC comprising at least 85% identity to the nucleic acids appearing as
 CC ABX11641 and ABX11642 or their complement, a naturally occurring variant
 CC of the kinases or their fragments. Also included are a non-human host
 CC cell containing the nucleic acids, an antibody specific for the proteins,
 CC identifying a compound which binds to the kinase (by contacting the
 CC kinase or a cell expressing the kinase with a test compound and
 CC determining whether the kinase binds to the test compound) and modulating
 CC the activity of kinase using the identified compound. The kinases and
 CC their encoding nucleic acids are useful as diagnostic and therapeutic
 CC agents for preventing a disease or condition associated with an aberrant
 CC or unwanted 59079 or 12599 activity in a subject, including
 CC cardiovascular diseases such as heart failure, and myocardial infarction;
 CC disorders involving blood vessels such as atherosclerosis, and Kaposi's
 CC sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,
 CC Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders
 CC such as cancer; and protein kinase disorders such as autoimmune
 CC disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,
 CC rheumatoid arthritis, and multiple sclerosis (many examples of diseases
 CC and disorders are included in the specification). The kinases, their
 CC encoding nucleic acids and antibodies are useful in screening assays,
 CC detection assays (e.g. forensic biology), and predictive medicine (e.g.
 CC diagnostic assays, prognostic assays, and monitoring clinical trials and
 CC pharmacogenomics). The kinases and their encoding nucleic acids are
 CC useful as query sequences to perform a search against public databases to
 CC identify other family members or related sequences. The present sequence
 CC represents kinase 59079
 XX
 SQ Sequence 2630 AA;

Query Match 33.2%; Score 13710; DB 6; Length 2630;
 Beat Local Similarity 100.0%; Pred. No. 0;
 Matches 2630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5339 MGAAYSSARLLVRGPDPEEKPSADVHEQLVPPRLERFTPKVKKGGSSITFSVKEGRP 5398
 Db 1 MGAAYSSARLLVRGPDPEEKPSADVHEQLVPPRLERFTPKVKKGGSSITFSVKEGRP 60
 QY 5399 VPTVHMLEEAEERGLWIGPDTPGYTVASSAQHSLVLLDVGROHQGTTCIASNAAGQA 5458
 Db 61 VPTVHMLEEAEERGLWIGPDTPGYTVASSAQHSLVLLDVGROHQGTTCIASNAAGQA 120
 QY 5459 LCSASLHVSGLPKVEEQSKVEALISTPLOGTTQAISAGLETAFADLGQRKKEPLAA 5518
 Db 121 LCSASLHVSGLPKVEEQSKVEALISTPLOGTTQAISAGLETAFADLGQRKKEPLAA 180
 QY 5519 KENLGHLSLAETGTEEFLOKLTSTQITEMVSAKITQAKLVQPGDSDSDSKTSAPRHR 5578
 Db 181 KENLGHLSLAETGTEEFLOKLTSTQITEMVSAKITQAKLVQPGDSDSDSKTSAPRHR 240
 QY 5579 SRPSSSIQESSSESDGDARGEIFDIYVVTADYLPGLAEQDAITLREGYVVLDAHPL 5638
 Db 241 SRPSSSIQESSSESDGDARGEIFDIYVVTADYLPGLAEQDAITLREGYVVLDAHPL 300
 QY 5639 RVLVTRKTPKSPRQGWVSPAYLDRRLKLSPEWGAAEAPFPFGPAVSEDEYKARLSSVI 5698
 Db 301 RVLVTRKTPKSPRQGWVSPAYLDRRLKLSPEWGAAEAPFPFGPAVSEDEYKARLSSVI 360
 QY 5699 QELLSSEQA FVEELQFLOSHLQHLERCPHPPIAVAGQKAVIFRNVDRIGRPHSSFLQEL 5758
 Db 361 QELLSSEQA FVEELQFLOSHLQHLERCPHPPIAVAGQKAVIFRNVDRIGRPHSSFLQEL 420
 QY 5759 QOCOTDDDDVAMCFIKNOAAFEQYLEFLVGRVQAESVVVSTAIQEFYKKYABEALLAGDPS 5818
 Db 421 QOCOTDDDDVAMCFIKNOAAFEQYLEFLVGRVQAESVVVSTAIQEFYKKYABEALLAGDPS 480

QY 5819 QPPPPLOHYLQBPVERVQRYQALLKELIRNKARNQNCALLQEQAYAVVSALPQRAENKL 5878
DB |||||
481 QPPPPLOHYLQBPVERVQRYQALLKELIRNKARNQNCALLQEQAYAVVSALPQRAENKL 540
QY 5879 HVSIMENYPTGLEALGEPIROGHFVWEGAPGARMPWKGNRHVFLFRNHLVICKPRRDS 5938
DB |||||
541 HVSIMENYPTGLEALGEPIROGHFVWEGAPGARMPWKGNRHVFLFRNHLVICKPRRDS 600
QY 5939 RDTVSYVFRNMWKLSSIDLNDQVGGDDRAFEVWQEREDSVRKYLLOQARTALIKSSWKE 5998
DB |||||
601 RDTVSYVFRNMWKLSSIDLNDQVGGDDRAFEVWQEREDSVRKYLLOQARTALIKSSWKE 660
QY 5999 ICGIQORLALPYWRPPDFEEELADCTAELGETVKLACRVGTGPKPVIWYKDGKAVQVDP 6058
DB |||||
661 ICGIQORLALPYWRPPDFEEELADCTAELGETVKLACRVGTGPKPVIWYKDGKAVQVDP 720
QY 6059 HHILIEDPDGSCALILDSITGVDSGOYMCFAASAAGNCSTLGKILVQVPPRPNVKVRASP 6118
DB |||||
721 HHILIEDPDGSCALILDSITGVDSGOYMCFAASAAGNCSTLGKILVQVPPRPNVKVRASP 780
QY 6119 FVEGEDAOPCTCTIEGAPYQIRWYKDGALLTTGNKFOTLSEPRSGLLVLVIRAASKEDLG 6178
DB |||||
781 FVEGEDAOPCTCTIEGAPYQIRWYKDGALLTTGNKFOTLSEPRSGLLVLVIRAASKEDLG 840
QY 6179 LYECELVNRLGARSASAEELRIQSPMLQAOEQCHREQLVAAVEDTTLERADQEVTSVLKRL 6238
DB |||||
841 LYECELVNRLGARSASAEELRIQSPMLQAOEQCHREQLVAAVEDTTLERADQEVTSVLKRL 900
QY 6239 LQPKAPGSGTDLTGPGCPRGAPALQETGSPVVTGTSAPVAPRVPVQPLLHGGPQOE 6298
DB |||||
901 LQPKAPGSGTDLTGPGCPRGAPALQETGSPVVTGTSAPVAPRVPVQPLLHGGPQOE 960
QY 6299 PEAIARAQEWTPVIRMEGAAMPAGCTGELLMDVHSHVVRRETTQRTYTYQADTHTARPPS 6358
DB |||||
961 PEAIARAQEWTPVIRMEGAAMPAGCTGELLMDVHSHVVRRETTQRTYTYQADTHTARPPS 1020
QY 6359 MQVTIEDVQAOTGGTAQFAEIIIEGDPQPSVTWYKDSVOLVDSTRLSQQOEGTYSVLVRH 6418
DB |||||
1021 MQVTIEDVQAOTGGTAQFAEIIIEGDPQPSVTWYKDSVOLVDSTRLSQQOEGTYSVLVRH 1080
QY 6419 VASKDAGVYTCIAQNTGGQVLCABELLVIGDNEPDSEKQHRKHLHGFYEVKEIGRGV 6478
DB |||||
1081 VASKDAGVYTCIAQNTGGQVLCABELLVIGDNEPDSEKQHRKHLHGFYEVKEIGRGV 1140
QY 6479 FGFVKRVQHGKNKILCAAKFIPLRSTRQAAYRERDILAAALSHPLVTLGLDQFETRKTLI 6538
DB |||||
1141 FGFVKRVQHGKNKILCAAKFIPLRSTRQAAYRERDILAAALSHPLVTLGLDQFETRKTLI 1200
QY 6539 LILELCSSEELLDRLYRGVWTEAEVKYIIOQLVEGLHYLHSHGVHLHDIKPSNITLMVHP 6598
DB |||||
1201 LILELCSSEELLDRLYRGVWTEAEVKYIIOQLVEGLHYLHSHGVHLHDIKPSNITLMVHP 1260
QY 6599 AREDIKICDPGFAQNTIPAELOFSGYSPFVSPBIIIOQNVPSEASDIWAMGVTSYLSLT 6658
DB |||||
1261 AREDIKICDPGFAQNTIPAELOFSGYSPFVSPBIIIOQNVPSEASDIWAMGVTSYLSLT 1320
QY 6659 CSSPPAGSDRATLNLVLEGRVSWSSPMAAHLSEDAKDFIKATLQAPQAPRPSAAQCLSH 6718
DB |||||
1321 CSSPPAGSDRATLNLVLEGRVSWSSPMAAHLSEDAKDFIKATLQAPQAPRPSAAQCLSH 1380
QY 6719 PWFLLKSMPEAEAHFINTKQLFLARSRWQSLMSYKSIILVWRSIPELLRGPDPSPSLGV 6778
DB |||||
1381 PWFLLKSMPEAEAHFINTKQLFLARSRWQSLMSYKSIILVWRSIPELLRGPDPSPSLGV 1440
QY 6779 ARHLCRDTGGSSSSSSSDNELAPPAKASIPPSPVTHSPLHLHPRGFILRPSASIPPEZAEA 6838
DB |||||
1441 ARHLCRDTGGSSSSSSSDNELAPPAKASIPPSPVTHSPLHLHPRGFILRPSASIPPEZAEA 1500
QY 6839 SERSTEAPAPPASPEGAPPAQGCVPRHVSIRSLFYHQAGESPBGHALAPGSRHHPAR 6898
DB |||||
1501 SERSTEAPAPPASPEGAPPAQGCVPRHVSIRSLFYHQAGESPBGHALAPGSRHHPAR 1560

QY 6899 RHLLKGGYIAGALPCLREPLMEHRVLEBEAAAREEQATLLAKAPFETALRLPASGTHLAP 6958
DB |||||
1561 RHLLKGGYIAGALPCLREPLMEHRVLEBEAAAREEQATLLAKAPFETALRLPASGTHLAP 1620
QY 6959 GHSHSLHDSSTPRPSSEACGEAQRLPASBSGAPITDMGHQPGSKOLPSTGGHPGTAQ 7018
DB |||||
1621 GHSHSLHDSSTPRPSSEACGEAQRLPASBSGAPITDMGHQPGSKOLPSTGGHPGTAQ 1680
QY 7019 PERPSPQWPQAPFCHPKQSGAPOEGCSPPHAPVAPCPGSPFPFGSCKEAPLVPSSPFL 7078
DB |||||
1681 PERPSPQWPQAPFCHPKQSGAPOEGCSPPHAPVAPCPGSPFPFGSCKEAPLVPSSPFL 1740
QY 7079 GQOQAPPAPAKASPPDLDSKMGPGDISLPGRPKPGPCSPSGSASQASSQVSSLRVSSQV 7138
DB |||||
1741 GQOQAPPAPAKASPPDLDSKMGPGDISLPGRPKPGPCSPSGSASQASSQVSSLRVSSQV 1800
QY 7139 GTEFGPSLDABGWTOEABEDLSDSTPTLQRPQOQATMRKFSLGGRGYAGVAGYGTFAFGG 7198
DB |||||
1801 GTEFGPSLDABGWTOEABEDLSDSTPTLQRPQOQATMRKFSLGGRGYAGVAGYGTFAFGG 1860
QY 7199 DAGMGLQGPWMAIANAVSQSEEEQEEARABESQSEBQEARAESPLQVYSARPVPBVG 7258
DB |||||
1861 DAGMGLQGPWMAIANAVSQSEEEQEEARABESQSEBQEARAESPLQVYSARPVPBVG 1920
QY 7259 RAPTRSSPEPTPWEIDIGQVSLVQIRDLSDGDAEADTISLDSI SEVDPAYINLSLDYDIKYL 7318
DB |||||
1921 RAPTRSSPEPTPWEIDIGQVSLVQIRDLSDGDAEADTISLDSI SEVDPAYINLSLDYDIKYL 1980
QY 7319 PFPEMI FRKVPKSAQPPPPSPMAEEELABPEPTWMPGELGPHAGLSEITSESDVDALL 7378
DB |||||
1981 PFPEMI FRKVPKSAQPPPPSPMAEEELABPEPTWMPGELGPHAGLSEITSESDVDALL 2040
QY 7379 AEAAGVRKRKWSRSLFHPGGRHLPLDREPAELGLRERVKASVEHISRIILKGRPEGLEK 7438
DB |||||
2041 AEAAGVRKRKWSRSLFHPGGRHLPLDREPAELGLRERVKASVEHISRIILKGRPEGLEK 2100
QY 7439 EGPPRKRPGLASFRLSGLKSWDRAPTFLRELSDTVVLGQSVTLACQVSAQAPAAQATWSK 7498
DB |||||
2101 EGPPRKRPGLASFRLSGLKSWDRAPTFLRELSDTVVLGQSVTLACQVSAQAPAAQATWSK 2160
QY 7499 DGAPLESSRVLISATLKNFOLLTILVVAEDLVYTCVSVSNALGTVTTTGVLRKAERPS 7558
DB |||||
2161 DGAPLESSRVLISATLKNFOLLTILVVAEDLVYTCVSVSNALGTVTTTGVLRKAERPS 2220
QY 7559 SSPCPDIEGVYADGVLLVWKPVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSKLSR 7618
DB |||||
2221 SSPCPDIEGVYADGVLLVWKPVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSKLSR 2280
QY 7619 GGTYTFRFTACVSKAGMGPYSSPSQVLLGGPSHLASEEESQGRSAQPLPSTKTFAFQTQI 7678
DB |||||
2281 GGTYTFRFTACVSKAGMGPYSSPSQVLLGGPSHLASEEESQGRSAQPLPSTKTFAFQTQI 2340
QY 7679 QRGFRFVVRQWKEKASGRALAAKIIPIHPKDKTAVLREYEAALKGLRHPHLAQLHAAVYLS 7738
DB |||||
2341 QRGFRFVVRQWKEKASGRALAAKIIPIHPKDKTAVLREYEAALKGLRHPHLAQLHAAVYLS 2400
QY 7739 RHLVLIILELCSGPPELLPCLAEASYSSEVKDYLMQMLSATQYLNHNOHILHLDRSENMI 7798
DB |||||
2401 RHLVLIILELCSGPPELLPCLAEASYSSEVKDYLMQMLSATQYLNHNOHILHLDRSENMI 2460
QY 7799 ITEYNLLKVVLDLNGNAQSLSQEKVLPDKFDYLETMAPELLEGQGVQPTDIIWAGVTAF 7858
DB |||||
2461 ITEYNLLKVVLDLNGNAQSLSQEKVLPDKFDYLETMAPELLEGQGVQPTDIIWAGVTAF 2520
QY 7859 IMLSAYFPVSESGARDLORGRLKGLVRLSRCYAGLSGGAVALRSTICAQPMWGPBCASS 7918
DB |||||
2521 IMLSAYFPVSESGARDLORGRLKGLVRLSRCYAGLSGGAVALRSTICAQPMWGPBCASS 2580
QY 7919 LQCPWLTEBGPACSRPAPVTFPTARLVFVRNKRKRALLYKRHNLAQVR 7968
DB |||||
2581 LQCPWLTEBGPACSRPAPVTFPTARLVFVRNKRKRALLYKRHNLAQVR 2630

PR 20-MAR-2001; 2001US-0277327P.
 PR 20-MAR-2001; 2001US-0277338P.
 PR 21-MAR-2001; 2001US-0277739P.
 PR 22-MAR-2001; 2001US-0277783P.
 PR 23-MAR-2001; 2001US-0278152P.
 PR 26-MAR-2001; 2001US-0278899P.
 PR 27-MAR-2001; 2001US-0278999P.
 PR 27-MAR-2001; 2001US-0279036P.
 PR 28-MAR-2001; 2001US-0279344P.
 PR 30-MAR-2001; 2001US-0279999P.
 PR 30-MAR-2001; 2001US-0280233P.
 PR 02-APR-2001; 2001US-0280802P.
 PR 02-APR-2001; 2001US-0280822P.
 PR 02-APR-2001; 2001US-0280900P.
 PR 04-APR-2001; 2001US-0281194P.
 PR 13-APR-2001; 2001US-0283675P.
 PR 13-APR-2001; 2001US-0287424P.
 PR 02-MAY-2001; 2001US-0288066P.
 PR 03-MAY-2001; 2001US-0288342P.
 PR 15-MAY-2001; 2001US-0291190P.
 PR 16-MAY-2001; 2001US-0291240P.
 PR 30-MAY-2001; 2001US-0294485P.
 PR 31-MAY-2001; 2001US-0294899P.
 PR 31-MAY-2001; 2001US-0294899P.
 PR 18-JUN-2001; 2001US-0299027P.
 PR 19-JUN-2001; 2001US-0299303P.
 PR 19-JUN-2001; 2001US-0299310P.
 PR 10-JUL-2001; 2001US-0304354P.
 PR 31-JUL-2001; 2001US-0309198P.
 PR 16-AUG-2001; 2001US-0312903P.
 PR 10-SEP-2001; 2001US-0318463P.
 PR 12-SEP-2001; 2001US-0318770P.
 PR 27-SEP-2001; 2001US-0325430P.
 PR 27-SEP-2001; 2001US-0325681P.
 PR 18-OCT-2001; 2001US-0330380P.
 PR 31-OCT-2001; 2001US-0335301P.
 PR 14-NOV-2001; 2001US-0332172P.
 PR 14-NOV-2001; 2001US-0332271P.
 PR 14-NOV-2001; 2001US-0332272P.
 PR 14-NOV-2001; 2001US-0333184P.
 PR 14-NOV-2001; 2001US-0333272P.
 PR 21-NOV-2001; 2001US-0332094P.
 PR 03-DEC-2001; 2001US-0337426P.
 PR 03-DEC-2001; 2001US-0338092P.
 PR 04-DEC-2001; 2001US-0337185P.
 PR 03-JAN-2002; 2002US-0345705P.
 PR 08-MAR-2002; 2002US-00093463.
 XX
 FA (CURA-) CURAGEN CORP.
 XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
 XX Boldog FL, Li L, Zerhusen BD, Tchernov VT, Gangolli EA, Vernet CM;
 PI Pena CE, Bargas CE, Liu X, Spytek KA, Gorman L, Spaderna SK;
 PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
 PI Taupier RJ, Padigar M, Shenoy SG, Kekuda R, Gusev VV, Pochart PF;
 PI Zhong M;
 XX WPI; 2002-732824/79.
 XX N-PSDB; ABV99362.
 XX
 XX New NOVX polypeptides and polynucleotides, useful for preventing,
 PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
 PT Alzheimer's disease, dyslipidemia, obesity, immune or hematopoietic
 PT disorders, and asthma.
 XX
 XX Claim 1; Page 137-138; 619pp; English.
 FS
 XX The present invention relates to new isolated proteins (NOVX) and their
 CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
 CC any number from 1 to 48. The NOVX proteins and coding sequences are
 CC useful in the manufacture of a medicament for treating a syndrome

CC associated with a human disease, preferably a NOVX-associated disorder.
 CC The NOVX coding sequences and proteins are useful for treating,
 CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
 CC obesity, infectious disease, anorexia, cancer-associated cachexia,
 CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
 CC disease, immune disorders, hematopoietic disorders, cardiovascular
 CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
 CC disturbances associated with obesity, metabolic syndrome X or wasting
 CC disorders associated with chronic diseases or various cancers. The NOVX
 CC coding sequences and proteins may also be used as targets for the
 CC identification of small molecules that modulate or inhibit e.g.
 CC neurogenesis, cell differentiation, cell proliferation, hematopoiesis,
 CC wound healing and angiogenesis, in gene therapy, in generation of
 CC antibodies that bind immunospecifically to NOVX substances for use in
 CC therapeutic or diagnostic methods
 XX
 SQ Sequence 4691 AA;
 Query Match 25 5%; Score 10519.5; DB 5; Length 4691;
 Best Local Similarity 51.1%; Pred. No. 0;
 Matches 2367; Conservative 118; Mismatches 342; Indels 1807; Gaps 66;
 QY 2595 MPLYNDSPFHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTARSLEVRVPVFLKALD 2654
 DB 1 MPLYNDSPFHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTARSLEVRVPVFLKALD 60
 QY 2655 DLSAERGTALQCEVSDPEAHVVRKDGVLGSPDKYDFLHTAGTGLVHVDVSPDAG 2714
 DB 61 DLSAERGTALQCEVSDPEAHVVRKDGVLGSPDKYDFLHTAGTGLVHVDVSPDAG 120
 QY 2715 LYTCHVGSEETRARVRVHDLHVGITKRLKTMVELEGSECSPECVLSHESADPAMTVGG 2774
 DB 121 LYTCHVGSEETRARVRVHDLHVGITKRLKTMVELEGSECSPECVLSHESADPAMTVGG 180
 QY 2775 KTVGSSRFQATROGRKYILVREAAPSDAGEVFSVRGLTSKASLIVREPAALIKPLE 2834
 DB 181 KTVGSSRFQATROGRKYILVREAAPSDAGEVFSVRGLTSKASLIVREPAALIKPLE 240
 QY 2835 DQWVAGEDVELRCELGRAGTFVHMLKDKRAIKRSQKYDVVCEGTMAVLIRGASLKDAG 2894
 DB 241 DQWVAGEDVELRCELGRAGTFVHMLKDKRAIKRSQKYDVVCEGTMAVLIRGASLKDAG 300
 QY 2895 EYTCEVASKSTASLHVEEKANCFTTELNLQVEKGTAVETCKTEHPAAVTVWRKGLLE 2954
 DB 301 EYTCEVASKSTASLHVEEKANCFTTELNLQVEKGTAVETCKTEHPAAVTVWRKGLLE 360
 QY 2955 LRASGKHQPSQEGTLRLTISALEKADSDTYTCIDIGQAQRAQLLVQRRVHIIELEDDV 3014
 DB 361 LRASGKHQPSQEGTLRLTISALEKADSDTYTCIDIGQAQRAQLLVQRRVHIIELEDDV 420
 QY 3015 DVQEGSSATFRCRISPANYEPVHWFLLDKTPLHANELNEIDAQPGGYHVLTLRQLAKDSG 3074
 DB 421 DVQEGSSATFRCRISPANYEPVHWFLLDKTPLHANELNEIDAQPGGYHVLTLRQLAKDSG 480
 QY 3075 TIYFEGDQASAAALRTERKPSVFSRELTDTATITEGEDTLVCTSTCDIIMCWTKDKGT 3134
 DB 481 TIYFEGDQASAAALRTERKPSVFSRELTDTATITEGEDTLVCTSTCDIIMCWTKDKGT 540
 QY 3135 LRGSARCOLSHEGHRQALLITGATLQDSGRYKCEGAGCACSSIVRVHARVPRFQEAALD 3194
 DB 541 LRGSARCOLSHEGHRQALLITGATLQDSGRYKCEGAGCACSSIVRVHARVPRFQEAALD 600
 QY 3195 EVLEGGATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRLNRLPQDSGRY 3254
 DB 601 EVLEGGATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRLNRLPQDSGRY 660
 QY 3255 SCSFGDQTTATLTVTALPAQFIKLRNKEATGATATLRCELSTKPAVEWRKSETLRD 3314
 DB 661 SCSFGDQTTATLTVTALPAQFIKLRNKEATGATATLRCELSTKPAVEWRKSETLRD 720
 QY 3315 GDYRCLQDQGMCELRGLAWVDAAEYSCVCGERTSASLTIRMPMPAHPTIRLNRHQSII 3374
 DB 721 GDYRCLQDQGMCELRGLAWVDAAEYSCVCGERTSASLTIRMPMPAHPTIRLNRHQSII 780

PR	08-MAR-2001;	2001US-0274194P.	DR	N-PSDB; ABV99363.
PR	08-MAR-2001;	2001US-0274281P.	XX	
PR	08-MAR-2001;	2001US-0274322P.	PT	New NOVX polypeptides and polynucleotides, useful for preventing,
PR	09-MAR-2001;	2001US-0274849P.	PT	diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
PR	12-MAR-2001;	2001US-0275235P.	PT	Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic
PR	13-MAR-2001;	2001US-0275578P.	PT	disorders, and asthma.
PR	13-MAR-2001;	2001US-0275579P.	XX	
PR	13-MAR-2001;	2001US-0275601P.	PS	Claim 1; Page 142-143; 619pp; English.
PR	14-MAR-2001;	2001US-0276000P.	XX	
PR	16-MAR-2001;	2001US-0276776P.	CC	The present invention relates to new isolated proteins (NOVX) and their
PR	19-MAR-2001;	2001US-0276994P.	CC	coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
PR	20-MAR-2001;	2001US-0277239P.	CC	any number from 1 to 48. The NOVX proteins and coding sequences are
PR	20-MAR-2001;	2001US-0277321P.	CC	useful in the manufacture of a medicament for treating a syndrome
PR	20-MAR-2001;	2001US-0277327P.	CC	associated with a human disease, preferably a NOVX-associated disorder.
PR	20-MAR-2001;	2001US-0277338P.	CC	The NOVX coding sequences and proteins are useful for treating,
PR	21-MAR-2001;	2001US-0277791P.	CC	preventing or diagnosing diseases such as metabolic disorders, diabetes,
PR	21-MAR-2001;	2001US-0277833P.	CC	obesity, infectious disease, anorexia, cancer-associated cachexia,
PR	23-MAR-2001;	2001US-0278152P.	CC	cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
PR	26-MAR-2001;	2001US-0278894P.	CC	disease, immune disorders, hematopoietic disorders, cardiovascular
PR	27-MAR-2001;	2001US-0278999P.	CC	disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
PR	27-MAR-2001;	2001US-0279036P.	CC	disturbances associated with obesity, metabolic syndrome X or wasting
PR	28-MAR-2001;	2001US-0279344P.	CC	disorders associated with chronic diseases or various cancers. The NOVX
PR	30-MAR-2001;	2001US-0279959P.	CC	coding sequences and proteins may also be used as targets for the
PR	30-MAR-2001;	2001US-0280233P.	CC	identification of small molecules that modulate or inhibit e.g.
PR	02-APR-2001;	2001US-0280802P.	CC	neurogenesis, cell differentiation, cell proliferation, haematopoiesis,
PR	02-APR-2001;	2001US-0280822P.	CC	wound healing and angiogenesis, in gene therapy, in generation of
PR	04-APR-2001;	2001US-0281194P.	CC	antibodies that bind immunospecifically to NOVX substances for use in
PR	13-APR-2001;	2001US-0283675P.	CC	therapeutic or diagnostic methods
PR	30-APR-2001;	2001US-0287424P.	XX	
PR	02-MAY-2001;	2001US-0288066P.	SQ	Sequence 4675 AA;
PR	03-MAY-2001;	2001US-0288342P.		
PR	03-MAY-2001;	2001US-0288528P.		
PR	15-MAY-2001;	2001US-0291190P.		
PR	16-MAY-2001;	2001US-0291099P.		
PR	16-MAY-2001;	2001US-0291240P.		
PR	30-MAY-2001;	2001US-0294485P.		
PR	31-MAY-2001;	2001US-0294889P.		
PR	31-MAY-2001;	2001US-0294899P.		
PR	18-JUN-2001;	2001US-0299027P.		
PR	19-JUN-2001;	2001US-0299303P.		
PR	19-JUN-2001;	2001US-0299310P.		
PR	10-JUL-2001;	2001US-0304354P.		
PR	31-JUL-2001;	2001US-0309198P.		
PR	16-AUG-2001;	2001US-0312903P.		
PR	10-SEP-2001;	2001US-0318462P.		
PR	12-SEP-2001;	2001US-0318770P.		
PR	27-SEP-2001;	2001US-0325430P.		
PR	27-SEP-2001;	2001US-0325681P.		
PR	18-OCT-2001;	2001US-0330380P.		
PR	31-OCT-2001;	2001US-0335301P.		
PR	14-NOV-2001;	2001US-0332172P.		
PR	14-NOV-2001;	2001US-0332271P.		
PR	14-NOV-2001;	2001US-0332272P.		
PR	14-NOV-2001;	2001US-0333184P.		
PR	14-NOV-2001;	2001US-0333272P.		
PR	21-NOV-2001;	2001US-0332094P.		
PR	03-DEC-2001;	2001US-0337426P.		
PR	03-DEC-2001;	2001US-0338092P.		
PR	04-DEC-2001;	2001US-0337185P.		
PR	03-JAN-2002;	2002US-0345705P.		
PR	08-MAR-2002;	2002US-00093463.		
XX				
PA	(CURA-) CURAGEN CORP.			
XX				
PI	Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;			
PI	Boldog FL, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Verniet CM;			
PI	Pena CEA, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;			
PI	Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;			
PI	Taupier RJ, Padigaru M, Shenoy SG, Kekuda R, Gusev VV, Pochart PF;			
PI	Zhong M;			
XX				
DR	WPI; 2002-732824/79.			

Db 541 LRGSARCOLSHEGHRQALLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQALXDL 600
Qy 3195 EVLEGGAATLRCVLSVAAPVWKCYNVLRPGDKYSLRQEGAMLELVVRNLRPODSGRY 3254
Db 601 EVLEGGAATLRCVLSVAAPVWKCYNVLRPGDKYSLRQEGAMLELVVRNLRPODSGRY 660
Qy 3255 SCSFGDQTTSATLTVTALPAQPIGKLRNKEATEGATATLRCELSKTAPVWVRKGSSETLRD 3314
Db 661 SCSFGDQTTSATLTVTALPAQPIGKLRNKEATEGATATLRCELSKTAPVWVRKGSSETLRD 720
Qy 3315 GDYRCLRDQGMACELOIRGLAMVDAEYSCVCGEERTSASLTIRPMPAHFIRGLRHQESI 3374
Db 721 GDYRCLRDQGMACELOIRGLAMVDAEYSCVCGEERTSASLTIRPMPAHFIRGLRHQESI 780
Qy 3375 EGATATLRCELSKAAPVWVRKGRRESLRDGRHSRLRODGAVALCELOICGLAVADAGYSVCV 3434
Db 781 EGATATLRCELSKAAPVWVRKGRRESLRDGRHSRLRODGAVALCELOICGLAVADAGYSVCV 840
Qy 3435 GEERTSATLTVKALPAKFTFEGLRNEAVEGATAMLCESKVAPVWVRKGPENLRDGRY 3494
Db 841 GEERTSATLTVKALPAKFTFEGLRNEAVEGATAMLCESKVAPVWVRKGPENLRDGRY 900
Qy 3495 ILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQAEAREGAT 3554
Db 901 ILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQAEAREGAT 960
Qy 3555 AVLOCELNSAAPVWVRKGSSETLRDGRYSLRQDGTCKELOIRGLAMADTGEYSVCVCGQER 3614
Db 961 AVLOCELNSAAPVWVRKGSSETLRDGRYSLRQDGTCKELOIRGLAMADTGEYSVCVCGQER 1020
Qy 3615 TSAMLTVRALPIKFTFEGLRNEATEGATAVLRCELSKMAPVWVRKGHETLRDGRHSLRQ 3674
Db 1021 TSAMLTVRALPIKFTFEGLRNEATEGATAVLRCELSKMAPVWVRKGHETLRDGRHSLRQ 1080
Qy 3675 DGARCELOIRGLVADAGEYLCMCKERTSAMLTVRAMPSPKFI EGLRNEEATEGDTATLW 3734
Db 1081 DGARCELOIRGLVADAGEYLCMCKERTSAMLTVRAMPSPKFI EGLRNEEATEGDTATLW 1140
Qy 3735 CELSKAAPVWVRKGHETLRDGRHSRLRODGRSCELQIRGLAVVDAGYSVCVCGQERTSAT 3794
Db 1141 CELSKAAPVWVRKGHETLRDGRHSRLRODGRSCELQIRGLAVVDAGYSVCVCGQERTSAT 1200
Qy 3795 LTVRALPARFIEDVKNQAEAREGATAVLOCELSKAAPVWVRKGSSETLRGGDRYSLRQDGT 3854
Db 1201 LTVRALPARFIEDVKNQAEAREGATAVLOCELSKAAPVWVRKGSSETLRGGDRYSLRQDGT 1260
Qy 3855 CELQIHGLSVADTGEYSVCVCGQERTSATLTVR ----- 3886
Db 1261 CELQIHGLSVADTGEYSVCVCGQERTSATLTVRALPARFTQDLKTKEASEGATATLQCELS 1320
Qy 3887 ----- 3886
Db 1321 KVAPVWVRKGPETLRDGRYSKQDGTGRCELOIHDLSVADAGEYSCMCGQERTSATLTVR 1380
Qy 3887 ----- 3886
Db 1381 ALPARFTFGLRNEEAMEGATATLQCELSKAAPVWVRKGLEALRDGDKYSLRQDGAVALCELO 1440
Qy 3887 ----- 3886
Db 1441 IHGLAMADNGYVSCVCGQERTSATLTVRALPARFIEDMRNQKATGATVTLQCKLRKAAP 1500
Qy 3887 ----- 3886
Db 1501 VEWKGPNTLKQDGRYSKQDGTSCELQIRGLVIADAGEYSCICEQERTSATLTVRALPA 1560
Qy 3887 ----- 3886
Db 1561 RFIEDVRNHEATEGATAVLOCELSKAAPVWVRKGSSETLRDGRYSLRQDGTGRCELOIRGL 1620
Qy 3887 ----- 3886
Db 1621 AVEDTGEYLCVCGQERTSATLTVRALPARFIDNMVNQAEAREGATATLHCELSKVAPVWVR 1680

Qy 3887 ----- 3886
Db 1681 KGPETLRDGRHSRLRODGRSCELQIRGLAVVDAGYSVCVCGQERTSATLTVRALPARFIE 1740
Qy 3887 ----- 3886
Db 1741 DVKNQAEAREGATAVLOCELSKAAPVWVRKGSSETLRGGDRYSLRQDGTGRCELOIHGLSVAD 1800
Qy 3887 ----- 3886
Db 1801 TGEYSCVCGQERTSATLTVRALPARFTQDLKTKEASEGATATLQCELSKVAPVWVRKGP 1860
Qy 3887 ----- 3886
Db 1861 TLRDGRYSKQDGTGRCELOIHDLSVADAGEYSCMCGQERTSATLTVRDCHTLHVMHPY 1920
Qy 3887 ----- 3886
Db 1921 FQLPGLLKEPETLIYIQIPSPVILFTFGLRNEEAMEGATATLQCELSKAAPVWVRKGL 1980
Qy 3887 ----- 3886
Db 1981 ALRDGDKYSLRQDGAVALCELOIHGLAMADNGYVSSLPARFIEDMRNQKATGATVTLQCKL 2040
Qy 3887 ----- 3886
Db 2041 RKAAPVWVRKGPNTLKQDGRYSKQDGTSCELQIRGLVIADAGEYSCICEQERTSATLTV 2100
Qy 3887 ----- 3886
Db 2101 RALPARFIEDVRNHEATEGATAVLOCELSKAAPVWVRKGSSETLRDGRYSLRQDGTGRCEL 2160
Qy 3887 ----- 3886
Db 2161 QIRGLVADTGEYLCVCGQERTSATLTVRALPARFIDNMVNQAEAREGATATLHCELSKVA 2220
Qy 3887 ----- 3886
Db 2221 PVWVRKGPETLRDGRHSRLRODGTGRCELOIRGLSVADAGEYSCVCGQERTSATLTVRALP 2280
Qy 3887 ----- 3886
Db 2281 AKFTKGLRNEATEGATAMLOCELSKVAPVWVRKGPETLRDGRYLNLRQDGTGRCELOIHG 2340
Qy 3887 ----- 3886
Db 2341 LSVADTGEYSVCVCGQERTSATLTVKAPQPVFPVFPVFPVFPVFPVFPVFPVFPVFPVFPV 3921
Qy 3922 WSKGGLQLOANGRRPREPLQGGCTABELVLQDLOREDTGEYTCCTCGSOATSATLTVTAAPVRF 3981
Db 2401 WSKGGLQLOANGRRPREPLQGGCTABELVLQDLOREDTGEYTCCTCGSOATSATLTVTAAPVRF 2460
Qy 3982 LRELQHOEVDDEGGTAHLCCESLRAGASVWVRKGSLOLPPCAKYQWVQDGAALVVRGVE 4041
Db 2461 LRELQHOEVDDEGGTAHLCCESLRAGASVWVRKGSLOLPPCAKYQWVQDGAALVVRGVE 2520
Qy 4042 QEDAGDYTCDTGHTQSMASLSVRVPKFKTRLOSLEOETGDIARLCCQLSDABESGAVVQ 4101
Db 2521 QEDAGDYTCDTGHTQSMASLSVRGGR -----GAA-- 2549
Qy 4102 WLKSGVELHAGPKYKEMRSQGTARELLIHQLBAKOTGEYACVTGGOKTAASLRVTEPEVTI 4161
Db 2550 -----CGPQVRDAAQGTARELLIHQLBAKOTGEYACVTGGOKTAASLRVTEPEVTI 2600
Qy 4162 VRGLVDAEVTDAEDVEFSCEVSRAGATGVQCLQGLPLQSNVEVEVAVRDRGRIHTRLKUG 4221
Db 2601 VRGLVDAEVTDAEDVEFSCEVSRAGATGVQCLQGLPLQSNVEVEVAVRDRGRIHTRLKUG 2660
Qy 4222 VTPEDAGTVSPHLGNHASSAQLTVRAPEVTITILEPLQDQVQLSEGQDASPCRLSRASGOBA 4281
Db 2661 VTPEDAGTVSPHLGNHASSAQLTVRAPEVTITILEPLQDQVQL ----- 2701

CC atherosclerosis, cirrhosis, hepatitis, paroxysmal nocturnal
CC haemoglobinuria, polycythaemia vera, psoriasis, primary
CC thrombocytopenia or cancer), developmental disorders (e.g. renal
CC tubular acidosis, anaemia or mental retardation), neurological disorders
CC (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), reproductive
CC disorders (e.g. infertility or a disruption in the menstrual cycle), or
CC autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma, autoimmune
CC thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus,
CC glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease,
CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
CC osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid
CC arthritis, Sjogren's syndrome, uveitis), or vital, bacterial, fungal,
XX parasitic, protozoal or helminthic infections
SQ Sequence 2328 AA;

Query Match 23.8%; Score 9834.5; DB 6; Length 2328;
Best Local Similarity 84.6%; Pred. No. 0;
Matches 1941; Conservative 76; Mismatches 213; Indels 65; Gaps 8;

QY 2096 MEVQLSHADVGSWTRDGLRFGQPTCHLAVRGPMTLTLGLRPEDSGLMVFKAEVHT 2155
DB 1 MEVQLSHADVEGSWTRDGLRFGQPTCHLAVRGPMTLTLGLRPEDSGLMVFKAEVHT 60

QY 2156 SARLVVTELPSFSPLQDVVTEKEKVTLECELSRPNDVRLKDGVELRAGKTMATAA 2215
DB 61 SARLVVTELPSFSPLQDVVTEKEKVTLECELSRPNDVRLKDGVELRAGKTMATAA 120

QY 2216 QGACHSLIYRCEPADQGVVYCDADHAQSSASVKVQ----- 2251
DB 121 QGACHSLIYRCEPADQGVVYCDADHAQSSASVKVQ----- 180

QY 2252 -----GRYTYLIYRVLAEADAGEIQFVAENAESRA 2281
DB 181 CEVSHDEVPQWFEGSKLRPTDNVRIQEGRTYTLIYRVLAEADAGEIQFVAENAESRA 240

QY 2282 QLRVKELPVTLPRLDKTAMEKHGVLGECQVSRASQVRMPKSGOELQPGPKYELVSDG 2341
DB 241 QLRVKELPVTLPRLDKTAMEKHGVLGECQVSRASQVRMPKSGOELQPGPKYELVSDG 300

QY 2342 LYRKLIISDVHAEDEDTYTCADGVKTSQAQFFVEEQSITIVRGLQDVTMEPAPAWFCE 2401
DB 301 LYRKLIISDVHAEDEDTYTCADGVKTSQAQFFVEEQSITIVRGLQDVTMEPAPAWFCE 360

QY 2402 TSIPSVRPPKLLGKTVLQAGNVGLEQGTVHRLMLRRTCTMTGPHVFTVGKSRSSAR 2461
DB 361 TSIPSVRPPKLLGKTVLQAGNVGLEQGTVHRLMLRRTCTMTGPHVFTVGKSRSSAR 420

QY 2462 LVVSDIPVLTPLPPTKGTRELOSVLSCDPRPAPKAVQWYKDDTPLSPSEKPKMSLEGQ 2521
DB 421 LVVSDIPVLTPLPPTKGTRELOSVLSCDPRPAPKAVQWYKDDTPLSPSEKPKMSLEGQ 480

QY 2522 MAELRILMPADAGVYRCQAGSAHSSTVTEAREVTVTGLODAEATEEGWASFSCEL 2581
DB 481 MAELRILMPADAGVYRCQAGSAHSSTVTEAREVTVTGLODAEATEEGWASFSCEL 540

QY 2582 SHEDVEWSLNGMPLYNDSFHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTSARLE 2641
DB 541 SHEDVEWSLNGMPLYNDSFHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTSARLE 600

QY 2642 VRVKPVVFLKALDDLASABERTLALQCEVSDPEAHVWRKDVQLGPSDKYDFLHTAGTR 2701
DB 601 VRVKPVVFLKALDDLASABERTLALQCEVSDPEAHVWRKDVQLGPSDKYDFLHTAGTR 660

QY 2702 GLVHVDVSPDAGLVTCVHGSSETRARVRVHDLHVGIITKRLKTMVELEGESCSFECVLSH 2761
DB 661 GLVHVDVSPDAGLVTCVHGSSETRARVRVHDLHVGIITKRLKTMVELEGESCSFECVLSH 720

QY 2762 ESADPAMWTVGGKTVGSSSRFQATROGRKYLLVVRREAPSDAGSWFVSVRGLTSKASLI 2821
DB 721 ESADPAMWTVGGKTVGSSSRFQATROGRKYLLVVRREAPSDAGSWFVSVRGLTSKASLI 780

QY 2822 VRERPAALIKPLEDQWVAPGEDVELURCELSSRAGTPVHMLKORAKRKQKDYDVCEGMA 2881

DB 781 VRERPAALIKPLEDQWVAPGEDVELURCELSSRAGTPVHMLKORAKRKQKDYDVCEGMA 840

QY 2882 MLVIRGASLKDAGEYTCVEASKSTASLHVBEKANKCFTEELTNLQVEBKGTAVFTCKTEH 2941

DB 841 MLVIRGASLKDAGEYTCVEASKSTASLHVBEKANKCFTEELTNLQVEBKGTAVFTCKTEH 900

QY 2942 PAATVWRKGLLELRASQKHPQSEGLTLRLTISALEKADSDTYTCDTGOASRAQLLVQ 3001

DB 901 PAATVWRKGLLELRASQKHPQSEGLTLRLTISALEKADSDTYTCDTGOASRAQLLVQ 960

QY 3002 GRRVHIITLEDVQVQEGSSATFCRISPNANYPVHWFDPKTPHLANELNEIDAOPGGVH 3061

DB 961 GRRVHIITLEDVQVQEGSSATFCRISPNANYPVHWFDPKTPHLANELNEIDAOPGGVH 1020

QY 3062 VLTLRQALKDQSGTYTPEAGDQASALRVTEKPSVFSRELTDTATITTEGEDTLVLCETST 3121

DB 1021 VLTLRQALKDQSGTYTPEAGDQASALRVTEKPSVFSRELTDTATITTEGEDTLVLCETST 1080

QY 3122 CDI PMCWTKDGTLRGSAARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSIVRVH 3181

DB 1081 CDI PMCWTKDGTLRGSAARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSIVRVH 1140

QY 3182 ARPVRFOALKDLEVLGGAATLRCVLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLEL 3241

DB 1141 ARPVRFOALKDLEVLGGAATLRCVLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLEL 1200

QY 3242 VVRNLRPDQSGRYSCSGFDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCLESKTA 3301

DB 1201 VVRNLRPDQSGRYSCSGFDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCLESKTA 1260

QY 3302 PVWRKGSSETLRDGDYCLRDQGMACELOIRGLAMVDAEYSCVCGEERTSASLTIRPMP 3361

DB 1261 PVWRKGSSETLRDGDYCLRDQGMACELOIRGLAMVDAEYSCVCGEERTSASLTIRPMP 1320

QY 3362 AHFTGLRHQESIEGATATLRCLESKAAPVWRKGRESLRDGDHSLRDQDGAVCLEQICG 3421

DB 1321 AHFTGLRHQESIEGATATLRCLESKAAPVWRKGRESLRDGDHSLRDQDGAVCLEQICG 1380

QY 3422 LAVADAGEYSCVCGEERTSATLTVKALPAKTEGLRNEEAEGATAMLCWELSKVAPVEM 3481

DB 1381 LAVADAGEYSCVCGEERTSATLTVKALPAKTEGLRNEEAEGATAMLCWELSKVAPVEM 1440

QY 3482 RKGPNLRDGDYCLRDQGMACELOIRGLAMVDAEYSCVCGEERTSASLTIRALPARFI 3541

DB 1441 RKGPNLRDGDYCLRDQGMACELOIRGLAMVDAEYSCVCGEERTSASLTIRALPARFI 1500

QY 3542 EDVKNQEAEGATAVLQCELNSAAPVWRKGSSETLRDGDYSLRDQGTCKELQIRGLAMA 3601

DB 1501 EDVKNQEAEGATAVLQCELNSAAPVWRKGSSETLRDGDYSLRDQGTCKELQIRGLAMA 1560

QY 3602 DTGEYSVCVCGEERTSASLTIVRALPIKFTTEGLRNEEAEGATAVLRCLESKVAPVEMWKGH 3661

DB 1561 DTGEYSVCVCGEERTSASLTIVRALPIKFTTEGLRNEEAEGATAVLRCLESKVAPVEMWKGH 1620

QY 3662 ETLRDGDHSLRDQGMACELOIRGLAMVDAEYLCMCKERTSASLTIVRAMPSKFI EGLR 3721

DB 1621 ETLRDGDHSLRDQGMACELOIRGLAMVDAEYLCMCKERTSASLTIVRAMPSKFI EGLR 1680

QY 3722 NEEATEGDTATLWCELSKAAAPVWRKGHETLRDGDHSLRDQGSRCLEQIRGLAVVDAGE 3781

DB 1681 NEEATEGDTATLWCELSKAAAPVWRKGHETLRDGDHSLRDQGSRCLEQIRGLAVVDAGE 1740

QY 3782 YSCVCGEERTSATLTIVRALPARFIEDVKNQEAEGATAVLQCELNSAAPVWRKGSSETLR 3841

DB 1741 YSCVCGEERTSATLTIVRALPARFIEDVKNQEAEGATAVLQCELNSAAPVWRKGSSETLR 1800

QY 3842 GGDYSLRDQGTCKELQIRGLAMVDAEYSCVCGEERTSASLTIVRAPOVFPREPQSLQA 3901

DB 1801 GGDYSLRDQGTCKELQIRGLAMVDAEYSCVCGEERTSASLTIVRAPOVFPREPQSLQA 1860

QY 3902 EEGSTATLQCELSSEPTATVWVSKGLQLOANGRREPRLQGGCTAEVLALDLOREDGEYTC 3961

DB 1861 EEGSTATLQCELSSEPTATVWVSKGLQLOANGRREPRLQGGCTAEVLALDLOREDGEYTC 1920

Db 1861 SEGATATLOCELSK-VAPVEWKGPETLDRDGRYSILKQDGTTRCELIQHDLVSADAGEVSC 1919
QY 3962 TCGSQATSLTVAAPVRFLELQHQEYDEGTAHLCCELSRAGASVWVRKGSLLQFPFC 4021
Db 1920 MCGQERTSAMLTVRALPARFTTEGLRNEEAMEGATATLOCELSKA-APVEWRKGLEALRDG 1978
QY 4022 AKYQWODGAARELLVRGVEQEDAGDYTCOTGHTQSMASLSVRVRPKTKRLQSLQEBT 4081
Db 1979 DRYSLRQDGAVCCELQIHGLAMADNGYSCVCGQERTSATLTVRALPARFIEDMRNQKATE 2038
QY 4082 GDIAUCCOLSAESGAVVQWLKEGVELHAGPKYEMRSOGATRELLIHOLEAKDGTGEVAC 4141
Db 2039 GATVTLQCKLRKA---APVEWRKGPNTLKDGRYSILKQDGTSCLEQIRGLVITADAGEVSC 2095
QY 4142 VTGGQKTAASLRVTEPEVTVIRGLVDAEVTADEDEFFSCEVSRAGATGVQWCLQGLPLQS 4201
Db 2096 ICEQERTSATLTVRALPARFIEDVRNHEATEGATAVLQCELSKAAP--VEWRKGSSETLRD 2153
QY 4202 NEVTEVAVNRDGRHILRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTTILEPIQDVOL 4261
Db 2154 GDRYSILR-QDGTTRCELIQIRGLAVEDTGEYLCVCGQERTSATLTVRALPARFIDNNTQGEA 2212
QY 4262 SEGQDASFQCLSRASGQEARWALGVPLQANEMNDITVEQCTLHLLTLHKVTLLEDAGTV 4321
Db 2213 REGATATLHCELSKAAPVE--WRKGRESLRDGRHSILR-QDGAVCCELQICGLAVADAGEY 2269
QY 4322 SFHVGTCSESAQLKV 4336
Db 2270 SCVCGEERTSATLTV 2284

RESULT 8

AAO15372
ID AAO15372 standard; protein; 1665 AA.

AC AAO15372;
XX
DT 19-SEP-2002 (first entry)
DE Human myosin light chain kinase subfamily-related kinase protein.
KW Human; gene therapy; chromosome 1; kinase protein;
KW myosin light chain kinase subfamily; kinase protein-mediated disease;
KW transgenic animal.

XX Homo sapiens.

XX WO200240683-A2.

XX 23-MAY-2002.

XX 22-OCT-2001; 2001WO-US032616.

XX 14-NOV-2000; 2000US-00711134.

PR 17-MAY-2001; 2001US-00858664.

XX (PEKE) PE CORP NY.

XX Wei M, Ketchum K, Di Francesco V, Beasley EM;
PI

XX WPI; 2002-500223/53.

DR N-PSDB; AAL43908, AAL43909.

XX New kinase proteins related to myosin light chain kinase subfamily and
PT encoding polynucleotide, useful for diagnosing, treating disease or
PT condition mediated by the kinase protein and for identifying modulators.

XX Claim 1; Fig 2; 96pp; English.

XX The invention comprises the amino acid and coding sequences (located on
CC chromosome 1) of a human kinase protein that is related to the myosin
CC light chain kinase subfamily. The human kinase DNA and protein sequences
CC of the invention are useful for identifying agents that modulate the

CC activity of the human kinase protein. Kinase-modulating agents are useful
CC for treating a disease or condition mediated by a human kinase protein.
CC The human kinase DNA sequences can be used to produce transgenic animals
CC which are useful for studying the function of kinase proteins and
CC identifying/evaluating modulators of kinase protein activity. The present
CC amino acid sequence represents the human kinase protein of the invention
XX
SQ Sequence 1665 AA;

Query Match 20.4%; Score 8423; DB 5; Length 1665;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6356 PPSQVITIEDYQAOTGGTAQFAEIIIEGDPQPSVTWYKDSVLQVDSVTRLSQQOEGTTSVLV 6415
Db 53 PPSQVITIEDYQAOTGGTAQFAEIIIEGDPQPSVTWYKDSVLQVDSVTRLSQQOEGTTSVLV 112
QY 6416 LRHVASKDAGVYVTCIAQNTGGQVLCKAELLVLGDNPDSEKQSHRRKLSHFYEVKEIG 6475
Db 113 LRHVASKDAGVYVTCIAQNTGGQVLCKAELLVLGDNPDSEKQSHRRKLSHFYEVKEIG 172
QY 6476 RGVFGFVKRVQHKNGKILCAAKFTPLRSRTQAQYRERDILAAALSHPLVTLGLDQFETRK 6535
Db 173 RGVFGFVKRVQHKNGKILCAAKFTPLRSRTQAQYRERDILAAALSHPLVTLGLDQFETRK 232
QY 6536 TLILILELCSSEELLDRLYRKGVVTEAEVKYIIQQLVEGLHYLHSHGVHLHDIKPSNLM 6595
Db 233 TLILILELCSSEELLDRLYRKGVVTEAEVKYIIQQLVEGLHYLHSHGVHLHDIKPSNLM 292
QY 6596 VHPAREDIKICDFGFAQNTIPAEILOFSQYSGPFPVSEIIOQNPVVSASDIWAMGVTSYL 6655
Db 293 VHPAREDIKICDFGFAQNTIPAEILOFSQYSGPFPVSEIIOQNPVVSASDIWAMGVTSYL 352
QY 6656 SLTCSPPAGESDRATLLNLEGRVSWSSPMAHLSEDADKDFIKATLQAPQAPPSAQC 6715
Db 353 SLTCSPPAGESDRATLLNLEGRVSWSSPMAHLSEDADKDFIKATLQAPQAPPSAQC 412
QY 6716 LSHPWFLKSPABEAHFINTKQLKFLLAGSRWORSLSYKSIILVMRSIPPELLRPPDPS 6775
Db 413 LSHPWFLKSPABEAHFINTKQLKFLLAGSRWORSLSYKSIILVMRSIPPELLRPPDPS 472
QY 6776 LGVARHLCRDTGGSS 6835
Db 473 LGVARHLCRDTGGSS 532
QY 6836 AEASERSTEAPAPASPEGAGPAAQCVPRHSVIRSLFYHQAGESPEHGALAPGSRHP 6895
Db 533 AEASERSTEAPAPASPEGAGPAAQCVPRHSVIRSLFYHQAGESPEHGALAPGSRHP 592
QY 6896 ARRRHLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSFETALRLPASGTH 6955
Db 593 ARRRHLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSFETALRLPASGTH 652
QY 6956 LAPGHSHLSLHSDSPSTPRPSSEACGEAQRILPSAPSGAPTRDMGHGPGSKQLSTGGHPG 7015
Db 653 LAPGHSHLSLHSDSPSTPRPSSEACGEAQRILPSAPSGAPTRDMGHGPGSKQLSTGGHPG 712
QY 7016 TAQPERPSDPSPMGQAPFCHPKQSGAPQSGCPHPAPAVPCPGSPPGCKEAPLVPSS 7075
Db 713 TAQPERPSDPSPMGQAPFCHPKQSGAPQSGCPHPAPAVPCPGSPPGCKEAPLVPSS 772
QY 7076 PFLGQPOAPPAPAKASPPPLDSKVGPGDILSLPGRPKPGCPSSPGSASQASSQVSSLRVGS 7135
Db 773 PFLGQPOAPPAPAKASPPPLDSKVGPGDILSLPGRPKPGCPSSPGSASQASSQVSSLRVGS 832
QY 7136 SOVGTPEGSLDAEGWTOEAEDLSDSTPTLORPOEQATWKEFSLGGGGYAGVAGYGTFA 7195
Db 833 SOVGTPEGSLDAEGWTOEAEDLSDSTPTLORPOEQATWKEFSLGGGGYAGVAGYGTFA 892
QY 7196 FGGDAGGMLQGGPMWARIWAVSQSEEEQEEAAREASQSEEQEAREAESPLPOVSARPVP 7255
Db 893 FGGDAGGMLQGGPMWARIWAVSQSEEEQEEAAREASQSEEQEAREAESPLPOVSARPVP 952

QY	6416	LRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGGDNEPDSEKQSHRRKHLHSFYVEKEEIG	6475	QY	7496	WSKGAPLESSRVLISATLKNFOLLTLVVVAEDLGVYTCVSNALGTVTTTGVLRKAE	7555
Db	113	LRHMASKDAGVYTCLAQNTGGQVLCKAELLVLGGDNEPDSEKQSHRRKHLHSFYVEKEEIG	172	Db	1193	WSKGAPLESSRVLISATLKNFOLLTLVVVAEDLGVYTCVSNALGTVTTTGVLRKAE	1252
QY	6476	RGVFGVKRVQHGKNGKILCAAKFIPILRSRTRAQAYRERDILAALSHPLVTGLDDQFETRK	6535	QY	7556	RPSSSPCPDICEVYADGVLLVWKPVESYGPVTYIVQCSLEGSSWTTLASDIFDCCYLTFSK	7615
Db	173	RGVFGVKRVQHGKNGKILCAAKFIPILRSRTRAQAYRERDILAALSHPLVTGLDDQFETRK	232	Db	1253	RPSSSPCPDICEVYADGVLLVWKPVESYGPVTYIVQCSLEGSSWTTLASDIFDCCYLTFSK	1312
QY	6536	TLILILELCSBELLDRLYRKGVWTEAEVKVYIQQLVEGLHYLHSHGVHLHDIKESNIIIM	6595	QY	7616	LSRGGTTFRTACVSKAGMGPSYSPQVLLGGPSHLASEESQGRSAQPLPSTKTFAQ	7675
Db	233	TLILILELCSBELLDRLYRKGVWTEAEVKVYIQQLVEGLHYLHSHGVHLHDIKESNIIIM	292	Db	1313	LSRGGTTFRTACVSKAGMGPSYSPQVLLGGPSHLASEESQGRSAQPLPSTKTFAQ	1372
QY	6596	VHPAREDIKIDFGFAQNTTPAELQFSQYSGSEFVSPEIIQONPVSEASDIWAMGVISYL	6655	QY	7676	TQIQRGFSVVRQWCKEASGRALAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQLHAAY	7735
Db	293	VHPAREDIKIDFGFAQNTTPAELQFSQYSGSEFVSPEIIQONPVSEASDIWAMGVISYL	352	Db	1373	TQIQRGFSVVRQWCKEASGRALAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQLHAAY	1432
QY	6656	SLTCSPPFAGESDRATLLNVLEGRVSWSPMAAHLSEDAKFIKATLQAPQAPSAQAQC	6715	QY	7736	LSPRHLVILIELCSGPBELLPCLAERASYSESEVKDYLMQMLSATQYILHNQHLHLDRSE	7795
Db	353	SLTCSPPFAGESDRATLLNVLEGRVSWSPMAAHLSEDAKFIKATLQAPQAPSAQAQC	412	Db	1433	LSPRHLVILIELCSGPBELLPCLAERASYSESEVKDYLMQMLSATQYILHNQHLHLDRSE	1492
QY	6716	LSHPWFLKSMPEAEAHFINTKQLKELLARSWRORSILMSYKSLVMRSIPELLRGPPDPS	6775	QY	7796	NMIITEYNLLKVVLDLGNASQISQSEKVLPSDKFKDYLETMAPELLEGOGAVPQTDIWAIGV	7855
Db	413	LSHPWFLKSMPEAEAHFINTKQLKELLARSWRORSILMSYKSLVMRSIPELLRGPPDPS	472	Db	1493	NMIITEYNLLKVVLDLGNASQISQSEKVLPSDKFKDYLETMAPELLEGOGAVPQTDIWAIGV	1552
QY	6776	LGVARHLCRDTGSSSSSSSSSSDNELAPFARAKSLPPSPVTHSPHLLHPRGFLRPSASLPEE	6835	QY	7856	TAFIMLSAEYPVPSSEGARDLQRLGRLKGLVRLSRCYAGLSGGAVAFILRSTLCAQPWGRPCA	7915
Db	473	LGVARHLCRDTGSSSSSSSSSSDNELAPFARAKSLPPSPVTHSPHLLHPRGFLRPSASLPEE	532	Db	1553	TAFIMLSAEYPVPSSEGARDLQRLGRLKGLVRLSRCYAGLSGGAVAFILRSTLCAQPWGRPCA	1612
QY	6836	ABASERSTAPAPPASPAGAPPAACGCVPRHSVIRSLFYHQAGSPSEHGALAPGSRRHP	6895	QY	7916	SSCLQCWLTEEGPACSRPAPVTPPTARLVRVFNREKRALLYKRHNLAQVR	7968
Db	533	ABASERSTAPAPPASPAGAPPAACGCVPRHSVIRSLFYHQAGSPSEHGALAPGSRRHP	592	Db	1613	SSCLQCWLTEEGPACSRPAPVTPPTARLVRVFNREKRALLYKRHNLAQVR	1665
QY	6896	ARRRHLKGGYTAGALPGLREPLMEHRVLEEAAREEQATLLAKAPSFETALRLPASGTH	6955	RESULT 10			
Db	593	ARRRHLKGGYTAGALPGLREPLMEHRVLEEAAREEQATLLAKAPSFETALRLPASGTH	652	AAB85504	ID	AAB85504	standard; protein; 1618 AA.
QY	6956	LAPGSHSLSEHDSPTPRPSSACGEAQLPSAPSGGAPIRDMGHPOGSKQLPSTGGHPG	7015	XX	AC	AAB85504;	
Db	653	LAPGSHSLSEHDSPTPRPSSACGEAQLPSAPSGGAPIRDMGHPOGSKQLPSTGGHPG	712	XX	DT	25-SEP-2001	(first entry)
QY	7016	TAQPERPSPDSWGPAPFCHPKQGSAPQEGGSPHPAVAPCPGPGFPFGSCKEAPLVPSS	7075	XX	DE	Human protein kinase SGK145.	
Db	713	TAQPERPSPDSWGPAPFCHPKQGSAPQEGGSPHPAVAPCPGPGFPFGSCKEAPLVPSS	772	XX	KW	Protein kinase; enzyme; cytostatic; nootropic; neuroprotective; human;	
QY	7076	PFLGQPPAPPAKASPPLDKMGPGDLSLGRPKPGPCSSPGSASQASSQVSSLRVGS	7135	XX	KW	antiparkinsonian; virucide; antibacterial; antifungal; antimigraine;	
Db	773	PFLGQPPAPPAKASPPLDKMGPGDLSLGRPKPGPCSSPGSASQASSQVSSLRVGS	832	XX	KW	analgesic; hypotensive; hypertensive; immunosuppressive; antiallergic;	
QY	7136	SOVGTEPGPSLDAEGWTQAEADLSDTPTLQRPQEQATMRKFSLGGRGYAGVAGYGTFA	7195	XX	KW	antipsoriatic; antirheumatic; antiarthritic; ophthalmological; anorectic;	
Db	833	SOVGTEPGPSLDAEGWTQAEADLSDTPTLQRPQEQATMRKFSLGGRGYAGVAGYGTFA	892	XX	KW	osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;	
QY	7196	FGDAGGMLGQPMWARIATAVVSQSEEEQEARAESQEQEARAESPLPVARSVP	7255	XX	OS	Homo sapiens.	
Db	893	FGDAGGMLGQPMWARIATAVVSQSEEEQEARAESQEQEARAESPLPVARSVP	952	XX	PN	WO200155356-A2.	
QY	7256	EVGRAPTRSSPEPTWEIDIGVSLVQIRDLSGDAEAADTISLDISEVDPAVLNLSLDYDI	7315	XX	PD	02-AUG-2001.	
Db	953	EVGRAPTRSSPEPTWEIDIGVSLVQIRDLSGDAEAADTISLDISEVDPAVLNLSLDYDI	1012	XX	PF	25-JAN-2001; 2001WO-US002337.	
QY	7316	KYLPEPFMI FRKVPKSAQPEPPSPMAEBELAEFPETWPFPGELGPHAGLSEEDVD	7375	XX	PR	25-JAN-2000; 2000US-0178078P.	
Db	1013	KYLPEPFMI FRKVPKSAQPEPPSPMAEBELAEFPETWPFPGELGPHAGLSEEDVD	1072	XX	PR	31-JAN-2000; 2000US-0179364P.	
QY	7376	ALLAEAAGVRKRWSSPSRSLFHPFGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEG	7435	XX	PR	17-FEB-2000; 2000US-0183173P.	
Db	1073	ALLAEAAGVRKRWSSPSRSLFHPFGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEG	1132	XX	PR	29-MAR-2000; 2000US-0190162P.	
QY	7436	LEKEGPPRKKPLGASFRISGLKSWMDRPTFLRELSDTVVLGQSVTLACQVSAQPAQAAT	7495	XX	PR	13-NOV-2000; 2000US-0247013P.	
Db	1133	LEKEGPPRKKPLGASFRISGLKSWMDRPTFLRELSDTVVLGQSVTLACQVSAQPAQAAT	1192	XX	PA	(SUGE-) SUGEN INC.	
				XX	PI	Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;	
				XX	DR	WPI; 2001-476202/51.	
				XX	DR	N-PSDB; AAH46904.	
				XX	PT	Kinase polypeptides useful for treating cancers, Alzheimer's disease,	

QY 7619 GCTYTRTACVSKAGMPYSSPSEQVLLGGPSHLASEEESQGRSAQPLPSTKTFQTOI 7678
 DB 1261 GCTYTRTACVSKAGMPYSSPSEQVLLGGPSHLASEEESQGRSAQPLPSTKTFQTOI 1320
 QY 7679 QRGFSVVRQCWEKASGRALAAKIIPIYHPKOKTAVLREYEAALKGLRHPHLAQHAAIYLS 7738
 DB 1321 QRGFSVVRQCWEKASGRALAAKIIPIYHPKOKTAVLREYEAALKGLRHPHLAQHAAIYLS 1380
 QY 7739 RHLVLILELCSGPELLCLAEERASVSESEVKDYILQMLSATQYLHNOHILHLDLSENMI 7798
 DB 1381 RHLVLILELCSGPELLCLAEERASVSESEVKDYILQMLSATQYLHNOHILHLDLSENMI 1440
 QY 7799 ITEYNLLKVDLGNAGLSQEKVLPDSKDFKDYLETMAPELLEGGQAVPOTDIWALGVTF 7858
 DB 1441 ITEYNLLKVDLGNAGLSQEKVLPDSKDFKDYLETMAPELLEGGQAVPOTDIWALGVTF 1500
 QY 7859 IMLSAEYPVSSGARDLQGLRKGLVRLSRCYAGLSGGAVAFRLSTLCAQPMGRPCASSC 7918
 DB 1501 IMLSAEYPVSSGARDLQGLRKGLVRLSRCYAGLSGGAVAFRLSTLCAQPMGRPCASSC 1560
 QY 7919 LQCPHLTEGPACSRPAPVTFPTARLRFVVRNREKRALLLYKRNLAQVR 7968
 DB 1561 LQCPHLTEGPACSRPAPVTFPTARLRFVVRNREKRALLLYKRNLAQVR 1610

RESULT 12 AAB30567

ID AAB30567 standard; protein; 1351 AA.

XX AC AAB30567;

XX DT 19-MAR-2001 (first entry)

XX DE Amino acid sequence of a human signal transduction polypeptide.

XX KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
 KW congestive heart failure; dilated congestive cardiomyopathy;
 KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
 KW mitral valve disease; aortic valve disease; tricuspid valve disease;
 KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
 KW atherosclerosis; cardiac tumour; microbial infection.

XX OS Homo sapiens.

XX PN WO200063381-A1.

XX PD 26-OCT-2000.

XX PF 11-APR-2000; 2000WO-US009488.

XX PR 16-APR-1999; 99US-0129553P.

XX PT (SCTO-) SCTOS INC.

XX PA Zeng W, Stanton L, Kong H;

XX PI WPI; 2001-007013/01.

XX DR N-PSDB; AAC62285.

XX PT Novel h19G5 polypeptides capable of regulating signal transduction and
 PT exhibiting kinase activity useful for identifying antibodies to treat
 PT cardiac diseases, and additional mediators of signal transduction.

XX PS Claim 1; Page 55-57; 81pp; English.

XX CC The present sequence represents a human protein with putative function in
 CC signal transduction. The polypeptide is designated H19G5. The protein is
 CC capable of regulating signal transduction and exhibits kinase activity.
 CC The H19G5 transcript is expressed in the heart. H19G5 polypeptides and
 CC polynucleotides are useful for preventing or treating a cardiac disease,
 CC such as congestive heart failure, dilated congestive cardiomyopathy,
 CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve
 CC disease, aortic valve disease or tricuspid valve disease, angina

CC pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial
 CC or renovascular hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours in humans. The polypeptide is also useful for detecting
 CC the expression of a protein capable of regulating signal transduction or
 CC the expression of a protein capable of acting as a donor or acceptor
 CC molecule of a phosphate group. The monoclonal antibodies can be used as
 CC probes for detecting discrete antigens expressed by tissue or cell
 CC samples, and therefore used in humans for localization and monitoring of
 CC microbial infection
 XX Sequence 1351 AA;

Query Match 17.1%; Score 7073; DB 4; Length 1351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1349; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6619 LQFQYSGPFPVPEIIQQNPVSEASDIWAMGVISYLSLTCTSSPPAGESDRATLLNVLEG 6678
 DB 2 VQFQYSGPFPVPEIIQQNPVSEASDIWAMGVISYLSLTCTSSPPAGESDRATLLNVLEG 61

QY 6679 RVSWSSPMAAHLSEDAKDFIKATLQAPQAPPSAAQCLSHWFLKSMPEAEAHFINTKQL 6738
 DB 62 RVSWSSPMAAHLSEDAKDFIKATLQAPQAPPSAAQCLSHWFLKSMPEAEAHFINTKQL 121

QY 6739 KFLIARSEWQBSLMSYKSIYVRSIPELLRGPPDPSIGVARHLCRDTCGSSSSSSSDN 6798
 DB 122 KFLIARSEWQBSLMSYKSIYVRSIPELLRGPPDPSIGVARHLCRDTCGSSSSSSSDN 181

QY 6799 ELAPFARAKSLPPSPVTHSPLLHPRGFLRPSASLPEEAEASERSTEAPAPASPGAGPP 6858
 DB 182 ELAPFARAKSLPPSPVTHSPLLHPRGFLRPSASLPEEAEASERSTEAPAPASPGAGPP 241

QY 6859 AAQCVPRHVSIRSLFYHQAGESPEHGALAPGSRHRHARRHLLKGGVIAGALPGLREPL 6918
 DB 242 AAQCVPRHVSIRSLFYHQAGESPEHGALAPGSRHRHARRHLLKGGVIAGALPGLREPL 301

QY 6919 MEHVRLEEARAREEQATLLAKAPFETALRLPASGTHLAPGSHLSHDSSTPRPSSEA 6978
 DB 302 MEHVRLEEARAREEQATLLAKAPFETALRLPASGTHLAPGSHLSHDSSTPRPSSEA 361

QY 6979 CGEAQRLPSAPSGGAPIRDMGHPOGSKQLPSTGGHPGTAQPERSPDPSFWQGPAPFCHPK 7038
 DB 362 CGEAQRLPSAPSGGAPIRDMGHPOGSKQLPSTGGHPGTAQPERSPDPSFWQGPAPFCHPK 421

QY 7039 QGSAPQEGCSPHPAVAPCPGSPFPFGSKCEAPLVPSSPFLGQAPAPAKASPPLD SKM 7098
 DB 422 QGSAPQEGCSPHPAVAPCPGSPFPFGSKCEAPLVPSSPFLGQAPAPAKASPPLD SKM 481

QY 7099 GPGDISLPGRPKGPCSSPGSASQASSSQVSLRVGSSQVGTGPGSLDAEGWTQEAEDL 7158
 DB 482 GPGDISLPGRPKGPCSSPGSASQASSSQVSLRVGSSQVGTGPGSLDAEGWTQEAEDL 541

QY 7159 SDSTPTLQRPQEQATMRKFSLGGRGVAGVAGYGTAFGGDAGGMLGQGPMMARLAWAVS 7218
 DB 542 SDSTPTLQRPQEQATMRKFSLGGRGVAGVAGYGTAFGGDAGGMLGQGPMMARLAWAVS 601

QY 7219 QSEEEQEEARAEASQSEEQEAEARAEPLQVARSVPVEVGRAPTRSSPEPTWEDIGQVS 7278
 DB 602 QSEEEQEEARAEASQSEEQEAEARAEPLQVARSVPVEVGRAPTRSSPEPTWEDIGQVS 661

QY 7279 LVQIRDLSGDAEAADTISLDISEVDPAVLNLSLDYDIKYLPEFMI FRKVPKSAQPEPPS 7338
 DB 662 LVQIRDLSGDAEAADTISLDISEVDPAVLNLSLDYDIKYLPEFMI FRKVPKSAQPEPPS 721

QY 7339 PMAEELAEFPETWMPGELGPHAGLEITTESEVDALLAEAAVGRKRKWSPPSRSLFH 7398
 DB 722 PMAEELAEFPETWMPGELGPHAGLEITTESEVDALLAEAAVGRKRKWSPPSRSLFH 781

QY 7399 FPGRHLPDLEPAELGLRERVKASVEHSIRILKGRPEGEKGGPPRKKPGLASFRUSGLKS 7458
 DB 782 FPGRHLPDLEPAELGLRERVKASVEHSIRILKGRPEGEKGGPPRKKPGLASFRUSGLKS 841

QY 7459 WDRAPTFLRELSDETIVLGQSVTLACQVSAQAPAAQATWSKOGAPLESSRVLISATLKNF 7518

Db 842 WDRAPTFLRELSDETIVLQSVTLACQVSAQPAQAATWSKDGAPLESSRVLISATLKNF 901
QY 7519 QLLTILVVAEDLGVVTCVSNALGVTTTGVLRKAERSSPCDDIGVYADGVLLVWK 7578
Db 902 QLLTILVVAEDLGVVTCVSNALGVTTTGVLRKAERSSPCDDIGVYADGVLLVWK 961
QY 7579 PVESYGPVTYIYQCSLEGGSWTTLASDIFDCCVLTSLSRGGTYTTPRTACVSKAGMPYS 7638
Db 962 PVESYGPVTYIYQCSLEGGSWTTLASDIFDCCVLTSLSRGGTYTTPRTACVSKAGMPYS 1021
QY 7639 SPSEQVLLGGPSHLASEESQGRSAQPLSTTKTFAPQTOIQGRFSVVRQCKEASGRAL 7698
Db 1022 SPSEQVLLGGPSHLASEESQGRSAQPLSTTKTFAPQTOIQGRFSVVRQCKEASGRAL 1081
QY 7699 AAKIIPYHPKDTAVLREYALKGLRHPLAOLHAAYLSPRHLVILELCSGPELLPCLA 7758
Db 1082 AAKIIPYHPKDTAVLREYALKGLRHPLAOLHAAYLSPRHLVILELCSGPELLPCLA 1141
QY 7759 ERASYESEVKDYLMQMSATQYLHNQHLHLDLRSENMIITEYNLLKVVDLGNQASLSQ 7818
Db 1142 ERASYESEVKDYLMQMSATQYLHNQHLHLDLRSENMIITEYNLLKVVDLGNQASLSQ 1201
QY 7819 EKVLPSDKPKYLETWAPPELLGGQAVPOTDIWAIGVTAFLMSAEYPVSSSEGARDLORG 7878
Db 1202 EKVLPSDKPKYLETWAPPELLGGQAVPOTDIWAIGVTAFLMSAEYPVSSSEGARDLORG 1261
QY 7879 LRKGLVRLSRCVAGLSGGAVAFRLSTLCAQWGRPCASSCLOCPWLITBEGPACSRPAPVT 7938
Db 1262 LRKGLVRLSRCVAGLSGGAVAFRLSTLCAQWGRPCASSCLOCPWLITBEGPACSRPAPVT 1321
QY 7939 FPTARLVFVRNREKRALLYKRHNLAQVR 7968
Db 1322 FPTARLVFVRNREKRALLYKRHNLAQVR 1351

RESULT 13
ADJ69186
ID ADJ69186 standard; protein; 1596 AA.
AC ADJ69186;
XX
DT 06-MAY-2004 (first entry)
XX Human heat mitochondrial protein as a therapeutic target Seqid992.
DE
XX
KW mitochondrial; human; screening assay; diabetes mellitus;
KW Huntington's disease; osteoarthritis;
KW Leber's hereditary optic neuropathy; LHON;
KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW myoclonic epilepsy tagged red fibre syndrome; MERRF; cancer;
KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW osteopathic; ophthalmological; cytostatic.

XX Homo sapiens.
OS
XX WO2003087768-A2.
XX
XX 23-OCT-2003.
XX
PF 04-APR-2003; 2003WO-US010870.
XX
XX 12-APR-2002; 2002US-0372843P.
PR 17-JUN-2002; 2002US-0389987P.
PR 20-SEP-2002; 2002US-0412418P.
XX
XX (MITO-) MITOKOR.
PA (BUCK-) BUCK INST AGE RES.
XX
PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI Warnock DE;
XX WPI; 2003-845369/78.

XX Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
XX with the disease.
PS Claim 1; SEQ ID NO 992; 180pp; English.
XX
CC This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, nootropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytostatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX
SQ Sequence 1596 AA;

Query Match 16.0%; Score 6619.5; DB 7; Length 1596;
Best Local Similarity 81.7%; Pred. No. 0;
Matches 1309; Conservative 70; Mismatches 203; Indels 21; Gaps 7;
QY 2726 RARVRHDLHVGIITKRLKTMVLESGSCFECVLSHESASDPAMWTGGKTVGSSSRFQA 2785
Db 1 RARVRHDLHVGIITKRLKTMVLESGSCFECVLSHESASDPAMWTGGKTVGSSSRFQA 60
QY 2786 TRQGRKYLIVVREAAAPSDAGEVWFSVRLGTSKASLIIVRERPAALIKPLEDQWAPGEDVE 2845
Db 61 TRQGRKYLIVVREAAAPSDAGEVWFSVRLGTSKASLIIVRERPAALIKPLEDQWAPGEDVE 120
QY 2846 LRCELSRAGTPVHLKDKRAIRKSQKYDVVCEGTMAMLVIRGASLKADAGEYTCVEASKS 2905
Db 121 LRCELSRAGTPVHLKDKRAIRKSQKYDVVCEGTMAMLVIRGASLKADAGEYTCVEASKS 180
QY 2906 TASLHVEKANCFTTELTNLQVEEKGTAVFTCKTEHPAATVTRKGLLEASGKHQPSQ 2965
Db 181 TASLHVEKANCFTTELTNLQVEEKGTAVFTCKTEHPAATVTRKGLLEASGKHQPSQ 240
QY 2966 EGLTLRLTISALEKADSDTYTCDIGQAQSAQLLVQGRRVHIIEDLEDVDVQEGSSATFR 3025
Db 241 EGLTLRLTISALEKADSDTYTCDIGQAQSAQLLVQGRRVHIIEDLEDVDVQEGSSATFR 300
QY 3026 CRISPANVPEVHWFLDKTPHLANELNEIDAQPGGYHVLTLRQALKDSGTYTFFAGDORA 3085
Db 301 CRISPANVPEVHWFLDKTPHLANELNEIDAQPGGYHVLTLRQALKDSGTYTFFAGDORA 360
QY 3086 SAALRVTEKPSVFSRELTDAITTEGEDTLVCETSTCDIIPMCWTKDGKTLRGSARCOLSH 3145
Db 361 SAALRVTEKPSVFSRELTDAITTEGEDTLVCETSTCDIIPVCWTKDGKTLRGSARCOLSH 420
QY 3146 EGHRAQLLITGATLQDSGRYKCEAGGACSSIVRVHARPVRFQALKDLVLEGGAATLR 3205
Db 421 EGHRAQLLITGATLQDSGRYKCEAGGACSSIVRVHARPVRFQALKDLVLEGGAATLR 480
QY 3206 CVLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPODSGRYSCSFGDQTTSA 3265
Db 481 CVLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPODSGRYSCSFGDQTTSA 540
QY 3266 TLTVTALPAQFIGLKRNEKATEGATLRLCELSKTAPEVWRKSETLRDGRYCLRDGA 3325
Db 541 TLTVTALPAQFIGLKRNEKATEGATLRLCELSKTAPEVWRKSETLRDGRYCLRDGA 600
QY 3326 MCELQIRGLAWDAAEYSCVCGEERTSASLTIRMPAHFTIGRLRHQESIEGATLRLCEL 3385
Db 601 MCELQIRGLAWDAAEYSCVCGEERTSASLTIRMPAHFTIGRLRHQESIEGATLRLCEL 660


```
QY 3386 SKAAPVWRKGRSRLDGRHSRLQDGAVALCELOICGLAVADAGEYSCVCGBERTSATITV 3445
Dd |||||
QY 661 SKAAPVWRKGRSRLDGRHSRLQDGAVALCELOICGLAVADAGEYSCVCGBERTSATITV 720
Dd |||||
QY 3446 KALPAKFTGLRNEEAVGATAMLCESLKVAPVWRKGPENLRDGRVILRQEGTRCEL 3505
Dd |||||
QY 721 KALPAKFTGLRNEEAVGATAMLCESLKVAPVWRKGPENLRDGRVILRQEGTRCEL 780
Dd |||||
QY 3506 QICGLAMADAGYLVCVCGQERTSATITRALPARFIEDVKNQEARREGATAVLQCELNSAA 3565
Dd |||||
QY 781 QICGLAMADAGYLVCVCGQERTSATITRALPARFIEDVKNQEARREGATAVLQCELNSAA 840
Dd |||||
QY 3566 PVWRKGSSETLRDGRYSRLQDGTCELOIRGLAMADTGEYSCVCGQERTSATITRALP 3625
Dd |||||
QY 841 PVWRKGSSETLRDGRYSRLQDGTCELOIRGLAMADTGEYSCVCGQERTSATITRALP 900
Dd |||||
QY 3626 IKFTGLRNEEATGATAVLRCELSKMAPVWRKGHETLRDGRHSRLQDGAVALCELOIRG 3685
Dd |||||
QY 901 IKFTGLRNEEATGATAVLRCELSKMAPVWRKGHETLRDGRHSRLQDGAVALCELOIRG 960
Dd |||||
QY 3686 LVAEDAGYLCMCGKERTSAMLTVRAMPKFTIEGLRNEEATGDTATLWCELSKAAPVEM 3745
Dd |||||
QY 961 LVAEDAGYLCMCGKERTSAMLTVRAMPKFTIEGLRNEEATGDTATLWCELSKAAPVEM 1020
Dd |||||
QY 3746 RKGHETLRDGRHSRLQDGRSRLQDGAVALCELOIRGLAVDAGEYSCVCGQERTSATITRALPARP 3805
Dd |||||
QY 1021 RKGHETLRDGRHSRLQDGRSRLQDGAVALCELOIRGLAVDAGEYSCVCGQERTSATITRALPARP 1080
Dd |||||
QY 3806 EDVKNQEARREGATAVLQCELSKAAPVWRKGSSETLRDGRYSRLQDGTCELOIRGLSVA 3865
Dd |||||
QY 1081 EDVKNQEARREGATAVLQCELSKAAPVWRKGSSETLRDGRYSRLQDGTCELOIRGLSVA 1140
Dd |||||
QY 3866 DTGEYSCVCGQERTSATITRALPARQVPRFQSLQAEESTATLQCELSSEPTATVWMSKG 3925
Dd |||||
QY 1141 DTGEYSCVCGQERTSATITRALPARFQDQTKASEGATATLQCELSK-VAPVEMKKG 1199
Dd |||||
QY 3926 GLOQANGRRPRLOGCTAELVLQDLOREDTGEYTCGSOATSATITVTAAPVRLREL 3985
Dd |||||
QY 1200 PETLRDGRYSRLQDGTCELOIRGLSVAADAGEYSCMCGQERTSATITRALPARFTEGL 1259
Dd |||||
QY 3986 QHOEYDEGTATLCELSELGRAGASVWRKGSLOLFPFCAKVMQVQDGAEEALLVRGVEQEDA 4045
Dd |||||
QY 1260 RNEEAMEGATATLQCELSKA-APVWRKGLERLQDGRKYSRLQDGAVALCELOIRGLAMADN 1318
Dd |||||
QY 4046 GDYTCDTGHTOSMASLSVRPRPKFTRQLQSLQEQETGDIARLCCQLSDAESGAVVQWLKE 4105
Dd |||||
QY 1319 GYVSCVCGQERTSATITRALPARFIEDVNRQKATEGATVTLQCKLRKA---APVEMRKG 1375
Dd |||||
QY 4106 GVHLAGPKYMRSGATRELLIHOLEAKDTGEYACVTGGQKTAASLRVTEPEVTIVRGL 4165
Dd |||||
QY 1376 PNTLRDGRYSRLQDGTCELOIRGLVADAGEYSCICEQERTSATITRALPARFIEDV 1435
Dd |||||
QY 4166 VDAEYVTADEDFEFSCEVRAGATGVQWCLQGLPLQSNVTEVAVRQDGRHITLRLKGVTEPE 4225
Dd |||||
QY 1436 RNEEAMEGATATLQCELSKAAP--VWRKGSSETLRDGRYSRLQDGTCELOIRGLAVE 1492
Dd |||||
QY 4226 DAGTVSFHLGNHASSAQITVRAPEVTILEPLQDVQLSEQQDASFCRLSRASGQARWAL 4285
Dd |||||
QY 1493 DTGEYLCVCGQERTSATITRALPARFIDNMNTNQEAREGATATLHCELSKAPVEM--WRK 1550
Dd |||||
QY 4286 GGVPLQANEMNDITVEQGTLLHLTLHKVTLEDAQTVSFHVGTC 4328
Dd |||||
QY 1551 GPETLRDGRHSRLQEN-----RLNPGGGCGGSELGSC 1582
Dd |||||
```

RESULT 14

AAB30570

ID AAB30570 standard; protein; 871 AA.

XX

AC AAB30570;

XX

DT 19-MAR-2001 (first entry)

XX

DE A splice variant of a signal transduction polypeptide.

XX

Signal transduction; H19G5; kinase; cardiac disease; angina pectoris; congestive heart failure; dilated congestive cardiomyopathy; hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension; mitral valve disease; aortic valve disease; tricuspid valve disease; myocardial infarction; cardiac arrhythmia; arteriosclerosis; atherosclerosis; cardiac tumour; microbial infection; splice variant.

OS

Homo sapiens.

XX

WO200063381-A1.

XX

26-OCT-2000.

PD

11-APR-2000; 2000WO-US009488.

XX

16-APR-1999; 99US-0129553P.

XX

(SCIO-) SCIOS INC.

PA

Zeng W, Stanton L, Kong H;

XX

WPI; 2001-007013/01.

DR

Novel h19G5 polypeptides capable of regulating signal transduction and exhibiting kinase activity useful for identifying antibodies to treat cardiac diseases, and additional mediators of signal transduction.

XX

Claim 1; Page 74-76; 81pp; English.

CC

The present sequence represents a splice variant of human in signal transduction polypeptide. The polypeptide is designated H19G5. The protein is capable of regulating signal transduction and exhibits kinase activity. The H19G5 transcript is expressed in the heart. H19G5 polypeptides and polynucleotides are useful for preventing or treating a cardiac disease, such as congestive heart failure, dilated congestive cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve disease, aortic valve disease or tricuspid valve disease, angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial or renovascular hypertension, arteriosclerosis, atherosclerosis and cardiac tumours in humans. The polypeptide is also useful for detecting the expression of a protein capable of regulating signal transduction or the expression of a protein capable of acting as a donor or acceptor molecule of a phosphate group. The monoclonal antibodies can be used as probes for detecting discrete antigens expressed by tissue or cell samples, and therefore used in humans for localization and monitoring of microbial infection.

XX

Sequence 871 AA;

SQ

Query Match

Best Local Similarity 11.0%; Score 4533; DB 4; Length 871;
Matches 871; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

7098 MGPDISLUPGRPKGPCSSPGSASQASSVSSLRVGSQVGTPEGPSLDAEGWQAEAD 7157
1 MGPDISLUPGRPKGPCSSPGSASQASSVSSLRVGSQVGTPEGPSLDAEGWQAEAD 60

Db

7158 LSDSTPTLQRQEQATMRKFSLLGGGGYAGVAGYCTFAFGDAGMLGGPMMARIWAV 7217
61 LSDSTPTLQRQEQATMRKFSLLGGGGYAGVAGYCTFAFGDAGMLGGPMMARIWAV 120

QY

7218 SQSEEEQEEARASQSEEQEARESPLPQVSARPVPEVGRAPTRSPPEPTWEDIGQV 7277
121 SQSEEEQEEARASQSEEQEARESPLPQVSARPVPEVGRAPTRSPPEPTWEDIGQV 180

Db

7278 SLVQIRDLSDGDAEAADTISLDISEVDPAVLNLSLDYIKLPPFPMIFRKVPKSAQPEPP 7337
181 SLVQIRDLSDGDAEAADTISLDISEVDPAVLNLSLDYIKLPPFPMIFRKVPKSAQPEPP 240

QY

7338 SPMAEEELAEPEPTWPGELGPHAGLEITEESEDVALLAEAVGRKRWKSSRSLSF 7397

Db 241 SPWAEELAEFPPTWPGELGPHAGLEITESEEDVDALLAAVGRKRKWSSESRSLF 300
QY 7398 HFFGRHLPLDEPAELGLRERVKASVEHISRLKGRPEGLEKEGPPRRKPKGLASFRLSGLK 7457
Db 301 HFFGRHLPLDEPAELGLRERVKASVEHISRLKGRPEGLEKEGPPRRKPKGLASFRLSGLK 360
QY 7458 SWDRAPTFRLSDETVVLGQSVTLACQVSAQAPAAQATWSKOGAPLESSSRVLISATLKN 7517
Db 361 SWDRAPTFRLSDETVVLGQSVTLACQVSAQAPAAQATWSKOGAPLESSSRVLISATLKN 420
QY 7518 FQLLTLLVVAEDLGVYTCVSNALGTVTGTVLKAERPSSPCPDIGENVADGVLLVW 7577
Db 421 FQLLTLLVVAEDLGVYTCVSNALGTVTGTVLKAERPSSPCPDIGENVADGVLLVW 480
QY 7578 KPVESYGPVTYIVQCSLEGSSWTLASDIFDCCYLTSLKSRGTYTFRACYSKAGMGYP 7637
Db 481 KPVESYGPVTYIVQCSLEGSSWTLASDIFDCCYLTSLKSRGTYTFRACYSKAGMGYP 540
QY 7638 SGPSEQVLLGGPSHLASEBESQGRSAQPLPSTKTFAFOTQIORGRFSVVVRQCEKASGRA 7697
Db 541 SGPSEQVLLGGPSHLASEBESQGRSAQPLPSTKTFAFOTQIORGRFSVVVRQCEKASGRA 600
QY 7698 LAAKIIPYHPKDKTAVLREVEALKGLRHPHLAQLHAAYLSPRHLVLIILELCSGPPELLPCL 7757
Db 601 LAAKIIPYHPKDKTAVLREVEALKGLRHPHLAQLHAAYLSPRHLVLIILELCSGPPELLPCL 660
QY 7758 AERASYSESEVKDYLMQWLSATQYLHNQHLHLDRSENMIITEYNLLKLVVDLGNQAQSL 7817
Db 661 AERASYSESEVKDYLMQWLSATQYLHNQHLHLDRSENMIITEYNLLKLVVDLGNQAQSL 720
QY 7818 QSKVLPSPDKFYLETMAPELLEGQAVPQTDIWAIGVTAFTMLSAEYVPSSEGARDLQR 7877
Db 721 QSKVLPSPDKFYLETMAPELLEGQAVPQTDIWAIGVTAFTMLSAEYVPSSEGARDLQR 780
QY 7878 GLRKGVLRLSRCVAGLSGAVAFRLSTLCAQPMWGPCASSCLOCPLWTEEGPACSRPAPV 7937
Db 781 GLRKGVLRLSRCVAGLSGAVAFRLSTLCAQPMWGPCASSCLOCPLWTEEGPACSRPAPV 840
QY 7938 TPTPTARLRFVVRNREKRALLYKRHNLAQVR 7968
Db 841 TPTPTARLRFVVRNREKRALLYKRHNLAQVR 871

RESULT 15
ID AAE16274
AC AAE16274; standard; protein; 871 AA.
XX AAE16274;
XX

DT 26-MAR-2002 (first entry)
DE Human kinase PKIN-20 protein.

KW Human; kinase; PKIN-20; cancer; leukaemia; adenocarcinoma; osteoporosis;
KW immune disorder; atherosclerosis; Crohn's disease; Hodgkin's disease;
KW Acquired Immune Deficiency Syndrome; AIDS; Addison's disease; anaemia;
KW allergy; asthma; adult respiratory distress syndrome; multiple sclerosis;
KW autoimmunity thyroiditis; bronchitis; diabetes mellitus; osteoarthritis;
KW Good pasture's syndrome; Graves' disease; pancreatitis; psoriasis;
KW rheumatoid arthritis; ulcerative colitis; cirrhosis; Cushing's syndrome;
KW hepatitis; hypothyroidism; cerebral palsy; cataract; angina pectoris;
KW cardiovascular disease; hypertension; vasculitis; myocarditis; obesity;
KW congestive heart failure; ischaemic heart disease; lung tumour; gout;
KW fatty liver; Niemann-Pick's disease; gene therapy.

OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 575..827
FT Domain /note= "Eukaryotic protein kinase domain"
FT 580..812
FT Domain /label= Protein_kinase_domain

XX WO200196547-A2.
XX 20-DEC-2001.
XX 14-JUN-2001; 2001WO-US019444.
XX 15-JUN-2000; 2000US-0212073P.
PR 23-JUN-2000; 2000US-0213467P.
PR 30-JUN-2000; 2000US-0215651P.
PR 07-JUL-2000; 2000US-0216605P.
PR 13-JUL-2000; 2000US-0218372P.
PR 25-AUG-2000; 2000US-0228056P.
XX (INCY-) INCYTE GENOMICS INC.
XX Yue H, Lal P, Bandman O, Borowsky ML, Au-Young J, Lu Y;
PI Gandhi AR, Tribouley CM, Wallia NK, Yao MG, Lu DAM, Greenwald SR;
PI Ramkumar J, Griffin JA, Kearney L, Burford N, Nguyen DB, Tang YT;
PI Baughn MR, He A, Thornton M, Hafalia A, Patterson C, Gururajan R;
PI Lo TP, Khan F, Recipon SA, Azimzai Y, Policky JL, Ding L;
PI Grether M, Elliott VS, Thangavelu K, Batra S, Ison CH;
XX WPI; 2002-090207/12.
DR N-PSDB; AAD26467.
XX
XX New polypeptides, useful for diagnosing, treating or preventing disorders
of growth and development, cardiovascular and lipid, and diseases such as
cancer, comprise human kinase polypeptides.
XX Claim 1; Page 164-165; 197pp; English.
XX The invention relates to human kinase PKIN proteins and their
corresponding cDNAs. A composition containing PKIN agonist is useful for
treating a disease or condition associated with decreased expression of
PKIN and a composition comprising PKIN antagonist is useful for treating
a disease or condition associated with overexpression of PKIN. The
disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma,
myeloma, sarcoma, teratocarcinoma, Hodgkin's disease); immune disorder
(Acquired Immune Deficiency Syndrome (AIDS), asthma, Addison's disease,
atherosclerosis, anaemia, allergies, adult respiratory distress syndrome,
autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes
mellitus, multiple sclerosis, Good pasture's syndrome, Graves' disease,
osteoarthritis, osteoporosis, pancreatitis, psoriasis, Reiter's syndrome,
rheumatoid arthritis, Sjogren's syndrome, uveitis, ulcerative colitis,
bacterial, parasitic, fungal, viral, protozoal and helminthic infections)
growth and development disorders (arteriosclerosis, cirrhosis, hepatitis,
Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts); cardio
vascular disease (arteriovenous fistula, hypertension, vasculitis,
aneurysms, congestive heart failure, angina pectoris, myocarditis,
ischaemic heart disease, chronic bronchitis, lung tumours); lipid
disorder (fatty liver, Fabry's disease, Niemann-Pick's disease,
CC hypocholesterolaemia, obesity). PKIN DNA is useful for assessing toxicity
of a test compound and in gene therapy. The present sequence is human
CC PKIN-20 protein
XX

SQ Sequence 871 AA;

Query Match 11.0%; Score 4529; DB 5; Length 871;
Best Local Similarity 99.9%; Pred. No. 6.2e-226;
Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 7098 MGFDISLPGKPKGPCSSPGSSQASSQSSVSLRVSSQVGTGPGSLDAEGWTQAEAD 7157
Db 1 MGFDISLPGKPKGPCSSPGSSQASSQSSVSLRVSSQVGTGPGSLDAEGWTQAEAD 60
QY 7158 LSDSTPLQRPQATMRKESLGGRGYAGVAGYCTFAFGDAGGMLGQGPMMARIWAV 7217
Db 61 LSDSTPLQRPQATMRKESLGGRGYAGVAGYCTFAFGDAGGMLGQGPMMARIWAV 120
QY 7218 SQSEEEQEAEARSQSEEQEAEARSPLQVSRPVPVEGRAPTRSSPPTWEDIGQV 7277
Db 121 SQSEEEQEAEARSQSEEQEAEARSPLQVSRPVPVEGRAPTRSSPPTWEDIGQV 180

```
QY 7278 SLVQIRDLSGDAEADTISLDISEVDPAVLNLSLDYDIKYLPFEFMI FRKVPKSAQPEPP 7337
Db      |||||||
QY 181 SLVQIRDLSGDAEADTISLDISEVDPAVLNLSLDYDIKYLPFEFMI FRKVPKSAQPEPP 240
Db      |||||||
QY 7338 SPMAEEELAEFPEPTWPGELGPHAGLEITESEDDALLAEAAVGRKRKWSPPSRSLF 7397
Db      |||||||
QY 241 SPMAEEELAEFPEPTWPGELGPHAGLEITESEDDALLAEAAVGRKRKWSPPSRSLF 300
Db      |||||||
QY 7398 HPPGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEGLEKGGPPRKPGLASFRLSGLK 7457
Db      |||||||
QY 301 HPPGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEGLEKGGPPRKPGLASFRLSGLK 360
Db      |||||||
QY 7458 SWDRAPTFLRELSDETIVLGQSVTLACQVSAQAATWSKDGAPLESSRVLISATLKN 7517
Db      |||||||
QY 361 SWDRAPTFLRELSDETIVLGQSVTLACQVSAQAATWSKDGAPLESSRVLISATLKN 420
Db      |||||||
QY 7518 FOLLTILVVVAEDLGVYTCVSNALGTVTTTGLRKAERPSSPCPDIGEVADGVLLVW 7577
Db      |||||||
QY 421 FOLLTILVVVAEDLGVYTCVSNALGTVTTTGLRKAERPSSPCPDIGEVADGVLLVW 480
Db      |||||||
QY 7578 KPVEYSGPVYIIVQCSLEGSWTTIASDIFDCCYLTSKLSRGTYTFRFTACVSKAGMGY 7637
Db      |||||||
QY 481 KPVEYSGPVYIIVQCSLEGSWTTIASDIFDCCYLTSKLSRGTYTFRFTACVSKAGMGY 540
Db      |||||||
QY 7638 SSPSEQVLLGGPSHLASEESQGRSAQPLPSTKTFAFQTIQGRFVSVRQCEKASGRA 7697
Db      |||||||
QY 541 SSPSEQVLLGGPSHLASEESQGRSAQPLPSTKTFAFQTIQGRFVSVRQCEKASGRA 600
Db      |||||||
QY 7698 LAAKIIPYHPKDKTAVLREYBALKGLRHPHLAQLHAAVLSRHLVLIILELCSGPPELLPCL 7757
Db      |||||||
QY 601 LAAKIIPYHPKDKTAVLREYBALKGLRHPHLAQLHAAVLSRHLVLIILELCSGPPELLPCL 660
Db      |||||||
QY 7758 AERASYSESEVKDYLLWQMLSATQYLLHNQHILHLDRSENMIITEYNLLKVVDLGNASLS 7817
Db      |||||||
QY 661 AERASYSESEVKDYLLWQMLSATQYLLHNQHILHLDRSENMIITEYNLLKVVDLGNASLS 720
Db      |||||||
QY 7818 QEKVLPSDKFDYLETMAPELLEGGQAVPQTDIWAIGVTAFIMLSAEYVVSSEGARDLQR 7877
Db      |||||||
QY 721 QEKVLPSDKFDYLETMAPELLEGGQAVPQTDIWAIGVTAFIMLSAEYVVSSEGARDLQR 780
Db      |||||||
QY 7878 GLRKLVLRLSRCYAGLSGGAFLRSTLCAQPMGRPCASSCLOCPWLTEEGPACSRPAPV 7937
Db      |||||||
QY 781 GLRKLVLRLSRCYAGLSGGAFLRSTLCAQPMGRPCASSCLOCPWLTEEGPACSRPAPV 840
Db      |||||||
QY 7938 TPTPTARLRFVVRNREKRALLYKRHNLAQVR 7968
Db      |||||||
QY 841 TPTPTARLRFVVRNREKRALLYKRHNLAQVR 871
Db      |||||||
```

Search completed: March 18, 2005, 17:42:51
Job time : 630 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2005, 17:26:12 ; Search time 146 Seconds
(without alignments)
5251.063 Million cell updates/sec

Title: US-10-077-130-5

Perfect score: 41273

Sequence: 1 MDQPFSGAPFLTRPKAFV.....RNREKRALLYKRHNLAQVR 7968

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3577.5	8.7	26926	1	I38344
2	3313	8.0	7962	2	I38346
3	2629	6.4	4162	2	T42633
4	2489	6.0	6642	2	T29757
5	2048	5.0	5198	2	T43290
6	2043.5	5.0	5175	2	T20992
7	1636.5	4.0	7160	2	T27935
8	1634	4.0	6839	2	S57242
9	1619	3.9	6831	2	A88852
10	1392	3.4	6658	2	T13931
11	1074	2.6	4391	2	A38096
12	1046.5	2.5	6805	2	S20901
13	1012	2.5	1906	1	S68235
14	963.5	2.3	1783	2	T34416
15	869	2.1	1323	2	PN0568
16	785.5	1.9	3488	2	T34418
17	767	1.9	3707	2	S18252
18	703.5	1.7	1398	2	T25568
19	645.5	1.6	1176	2	JN0583
20	622	1.5	1147	2	A59307
21	571	1.4	1274	2	S55050
22	532	1.3	1694	2	S50065
23	528.5	1.3	1132	2	A35089
24	520	1.3	1142	2	S36846
25	519	1.3	3375	2	T19821
26	501.5	1.2	608	2	A35021
27	480.5	1.2	610	2	A28798
28	472	1.1	1896	2	T08851
29	465.5	1.1	2541	2	T29340

30	463.5	1.1	451	2	S49128
31	459	1.1	1612	2	T30805
32	456.5	1.1	1398	2	T13741
33	456.5	1.1	1465	2	S43529
34	456	1.1	1423	1	I37275
35	456	1.1	1651	2	T14160
36	451.5	1.1	1051	1	JW0051
37	451.5	1.1	1451	2	S42167
38	451	1.1	371	2	JC7733
39	451	1.1	1123	2	S36846
40	445	1.1	1734	2	A54602
41	441.5	1.1	374	1	S50193
42	439.5	1.1	1450	2	A44027
43	427.5	1.0	1435	2	T32930
44	427	1.0	370	1	S57347
45	427	1.0	1138	2	S24614

ALIGNMENTS

RESULT 1

I38344
titin, cardiac muscle [validated] - human
N:Alternate names: connectin
N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004
C:Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R:Label: S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38344
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EH
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
R:Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:91017424; PID:91017425
R:Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix
A:Reference number: I38345; MUID:95119041; PMID:7819249
A:Accession: I38345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:G602579; PIDN:CAAS8243.1; PID:G602580
A:Note: conformation and properties are reported for a synthetic peptide corresponding
R:Label: S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380; PMID:1582406
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: I3597-14200, I', 14202-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:G37192; PIDN:CAA45939.1; PID:G37193
A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, S', 16384-16756, P', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:G37190; PIDN:CAA45940.1; PID:G37191
A:Accession: S20899
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'O', 22455-22480, 'TR', 22483-2
A:Cross-references: EMBL:X64697; NID:G37190; PIDN:CAA45938.1; PID:G37195
R:Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labelt, S.
J. Mol. Biol. 256, 556-563, 1996
A:Title: Genomic organization of M line titin and its tissue-specific expression in two
A:Reference number: S63665; MUID:96177761; PMID:8604138
A:Accession: S63665
A:Status: nucleic acid sequence not shown

A:Molecule type: DNA
A:Residues: 26729-26825 <KOL>
A:Cross-references: EMBL:X92412; NID:g1236761
R:Gautel, M.; Leonard, K.; Labelit, S.
EMBO J. 12, 3827-3834, 1993
A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiat
A:Reference number: S37393; MUID:94008990; PMID:8404852
A:Accession: S37393
A:Molecule type: mRNA
A:Residues: 26831-26926 <GAU>
R:Improra, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A66736; PDB:1TIT
A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R:Fuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A66201; PDB:INCT
A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q32
C:Function:
A:Description: structural protein forming filaments in striated muscle
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro
C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein
F:24752-25008/Domain: protein kinase homology <KIN>
F:84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4068,
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
tatus predicted
F:16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18
,21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248
F:26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 8.7%; Score 3577.5; DB 1; Length 26926;
Best Local Similarity 20.2%; Pred. No. 3.6e-95;
Matches 1929; Conservative 1288; Mismatches 3662; Indels 2663; Gaps 320;

QY	4	POFGAPFLTRPKAFVSVGKDATLSQIVGNPTPOVSWEKDQOPVTAGARFLA---	Q	60
Db	1030	PGEPAAPFITKPVVQKLVGGSVVFGQVCGNPKPHVYWKSGVPLTITGYRYKSYNKQ		1089
QY	61	DGDLVRLTLDLALGDSGYVCRARNATGEAFAAVGL-----	QVD	100
Db	1090	TGEC-KLVISMTFADDAGETYIVVRNKGGETSASASLLEADYELLMKSQOEMLYQTQVT		1148
QY	101	A---EACACAEQAPFL-----		113
Db	1149	AFVQPEVGETAPGFVYSEYEKEQALIRKKMAKDTVVVRTTVEBOEFHISFEERL		1208
QY	114	-----LRPTSIRVREGSEATFRQV		133
Db	1209	IKEIEYRIKTTLEELLEDGEKMAVDISEAEVSGFDLRKNYRILEGNGVTFHCKM		1268
QY	134	GGSPAPVSWKDGRLGPDGPRVVRBELGEASA-LRIRARPDGGTYEVRAENPLGA		192
Db	1269	SGYPLFKIAWKDGKRI--KHGERYQMDFLQDGRASLRIPVVLPEDEGIYTAFAASNIKGN		1326
QY	193	ASAAALVVDSDAADTASR--PGTSTAALLAHLQR---REAMRAEGAPASPP-----		240
Db	1327	AICSKLVPEAPALGAPTYITPLEPVSKIRLSLSPRSVSRSPIRMSPARMSPARMSPARM		1386
QY	241	-----STGTR-----TCTVTTEGKHARLSYVYTGSPKPTETVWKK		273
Db	1387	SPARMSPGRRLEETDESQLERLYKPVFLKPVSKLEGANCRCPLKVVGRPMETFWFH		1446
QY	274	DGQLVTEGRRH--VVYEDAQENFVLKILFCQSDRGLYTCFASNLVGGTYSVLVV---		328
Db	1447	DGQIVVNDYTHKWKVIGDGTSLI--IVPATPDSGSEWTVVAQNRAGRSSISVILTVEAV		1504
QY	329	-----		328

Db	1505	EHQVKPMFEVKLKNVNIKEGSRLEMKVRATGNPNPDIVMLKNSDIIVPHKYPKIRIEGTK		1564
QY	329	-----R		329
Db	1565	GEAALKIDSTVSQDSAWYTATAINKAGRDTRCKNVNVEFAEPEPEKLLIPROTYPAK		1624
QY	330	EPAPV-----FKKLODLEVREKESATFLC-----VP		357
Db	1625	EIAAPELEPLHLYRQEQWEEDLYDKEQKQPFKKKLTSLRKRFGPAHFCECLTIS		1684
QY	358	QPSTEAAWFKBETRI-----WASAKYGI-----BEEGTERLIT		390
Db	1685	DPTVWVEWLHDGKPLEAANRLRMINFEGYCSLDGVAYSRDSGIITCRATNKYGDHTSA		1744
QY	391	VRNSADDDAVVICETPEGSR-----TVASLAVQGNL-----LRKLPRKT		430
Db	1745	TLIVKDEKSLVESQLPEGRKGLORIIELEMAHEGALTGVTTQKEKQKQKEDIVLYPBPV		1804
QY	431	AVRVGDTAMF-CVELAVPVGPVHWRNQBEEVAVAGRVAISAEGRTRHTLTIISQCCLLEDVQ		489
Db	1805	RVLEGETARFCRVTPQPKVNVNLGQLIRKSRFRVRYDGI-HYLDIVDCKSYDTGE		1863
QY	490	VAFVAG-----DCQTSTRFCVSAPRKPLQP-----PVD		518
Db	1864	KVTAENPEGVIEHKVKLEIQOREDFRSVLRRAPPEPPEFHVHPEPKLQFEVQKVDREPD		1923
QY	519	-----PVVKARMESSVILSWSPPHGE-----RPVTIDGY-----LVEKKKLGVTW		560
Db	1924	TTEYKEVVKLKRARIIEHKVPESEBLRSKFRTREGYVEAITAVELSKRKDESYEE		1983
QY	561	I-----RCHAEAWATPBLTVAD-----VABEGNFQFRVSALNSFGQSPYLEPFGTVHLAPKL		613
Db	1984	LLRKTCKDELLHW--TKELTBEEKALAEBGKITPTTFKPKDKIELSPSME-----APKI		2034
QY	614	AVRPLKAVQAVEGEVTFPSVLDLTVASAG-----EWFLDQALKASS-VYEIHCDRTRH		666
Db	2035	PERIQSTVIG--QGSDAHFRVR-----VGKPDPECEYKNGVKIERSDRIYVWPEDNVC		2088
QY	667	TLTIREVPASLHGAKLKF-----VANGIESSTRMEVRAAPGLTANKPPAAAREVLARLHE		722
Db	2089	ELVIRDVTAB-DSASINVKAINIAGETSSHAFLVQAKLITFTQ-----ELQDVVAKEK		2143
QY	723	E-AQLLAELSDQAAAVTWLKDRTLSFGPKYEVQASAGRVLVLRDVARDDAGLYECV--		779
Db	2144	TMATFECETSEFPVKVWKYKDGMEVHEGDKYRMHSDRKVHFLSLITLITDSDAEDYSCVLV		2203
QY	780	-SRGRIAYOLSVQGLARFLHKOMAGSCVDAVAGGPAQFECETSEAHVHVHVKDGMELG		838
Db	2204	EDENVKTTAKLIVAGAVVVFVKELQD--IEVPESYSGELECTVSPENIEGKWHNDVELK		2261
QY	839	HSGERFLQEDVGRHRLVAATVTRQDEGTYSRVEDSDVDFRLRVSBPKVVFVFAKEQLARR		898
Db	2262	SNKYTITSRG-RQNLTVKDVTKEDQGEYSFVIDGKTKTKLKM-KPRPAILIQLGSDQ		2319
QY	899	KLQAEAGASATLCEVAQAQTEVTWYKDGKLLSSSKSKVCMEATGCTRRLLVQQAQADAG		958
Db	2320	KVC--EGDIVQLEVKVLSVESVEGVMKDGQEVPSDRVHVIDKQSHMLLIETDMTKEDAG		2377
QY	959	EYS-----CEAGGQRLSFHLVKEPKVVFVAKDQVAHSEVQAEAGANATLCEVAQAQ-		1010
Db	2378	NYSTPTPALGLSTSGRVSVSDVITPL-----KDVNVIEGKFAVLECKVSPVDV		2427
QY	1011	AEVNVYKDGKLLSSKLVHVEAKGRRRLVQQAQKTDAGDYSCEARGQVSRFLHITTEP		1070
Db	2428	TSVKWYLNDSQIKPDDRVQAI VKGTQRVLINRTHASDEGPKYKLVGRVETNCNLSVEKI		2487
QY	1071	KMPFAKEQSVHNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLLSSSKSKVCMEVKGCTRR		1130
Db	2488	KIIRGLRDLTCTCTEQ-----NVVFEVLSHSGIDVLMNPKDKBIKPSKSKYKIEAHGRIYK		2542
QY	1131	LVLPOAGKADAGYSCEAGGQVSRFHLHITPEKGVFAKEQSVHNEVQAEAGATGATMLSCV		1190
Db	2543	LTVLNMKDKDEGKTYTFAGENMTSGKL--TVAGASISK--PLTDQTVAES-QEAVFCEV		2597


```
Db      8709 VVTRKMKCLNWSDEDDGSGSEITGFI--IERKDKM----- 8743
Qy      6516 LAALSHPLVTGLLDQDPQTRKTLILILELCSSEELDLRYKGVVTEAEVKVYIQOLVREL 6575
Db      8744 -----HTWRQPIETERS-KCDITGLLEGQ 8766
Qy      6576 HYLHSHGVL-----HLDIKPSNLMVHPAREDIKICDFGAQNTIPABELQFSQYGS 6626
Db      8767 EY--KFRVIAKNKFGCGPPVEIGP--LLAVDP-----LGP 8797
Qy      6627 PRFVSPETIIQQNPVSEASDIWAGVISLSLTCSSPPFAGESDRATL-LNVLEGRVSWSSP 6685
Db      8798 P--TSPE-----RLAVT-----ERQSRITITLDWKEPRNGGSP 8828
Qy      6686 MAHLSE---DAKOFIKATLQAPQARPSAAQCLSHPWFLSKMAEAEAHFINTQKLKFL 6741
Db      8829 IGYIIEKRHRDKDPF-----ERVNKLCLPTTSPLENLDEHQMYEPRVK----- 8873
Qy      6742 LARSRWQRLMSYKSGILVMRSIPELLRGPDPSPSLGVARHLCRDTGGSSSSSSSDNELA 6801
Db      8874 -----AVNEIGSEPSLP-LNVV-----IQDDEVP 8897
Qy      6802 PPARAK-----SLP---PSPVTHSP---LHPRGFLRPSASL---PERAE 6837
Db      8898 PTIKRLSVRGDTIKVKAGEPVHIADVTGLPMPKIEWSKNETVIEKPTDALQITKEEVS 8957
Qy      6838 ASERSTEAPAPASPEGAPPAQCVPRHSHVIRSLFYHQAGESPEHGALAPGSRHRPAR 6897
Db      8958 RSEAKTELIPKAVREDKGTVTYVTSNRLGVSFRNVHVEYDR-----PSPPRNLAV 9009
Qy      6898 RRHLKGGYIAGALP---GLREPLMEHRVLEBEAAREQA-----TLAK-----A 6940
Db      9010 TDIKAESCVLTWADPLDNGGSE--ITHVIDKRDASRKKAEEVNTTAVEKRYGIWKL 9067
Qy      6941 PSFETALRLPAGTHLAPGHSHLSLHSDSPTRPSPSEACGEAQRPLSPASPGGAPTRDMGH 7000
Db      9068 PNGOYEFVRVAVNKY---GISDECKSDKVIQDP-----YRLPGPP--GKP-KVLAR 9113
Qy      7001 POGSKQLPST---GSHPGTAQ--PERPSPDSP-WGQAPAPFCHPKQG-----SAPO--E 7045
Db      9114 TKGSMVSWTPPLDNGGSPITGYWLEKEEGSPYWSRVSRAPITKVGLKGVFNVPRLLE 9173
Qy      7046 GCS-PHPAVAPCPGSPFGGCKEAPLVFPSS-FLGQAPAPAPAKASPPLDKMGPGDI 7103
Db      9174 GVKYQFRAMAINAAGIGPSE-----PSDDEVAGDPIFPFGP-PSCPVEVKDT-KSSI 9224
Qy      7104 SLPGPKPGPCSSPSASQASSQVSSLRVGSQVGTGPGPSLDAGM--TOEABDLSDS 7161
Db      9225 SLGMRP---PAKDG---SPIKGYIVEMQEEGT-----TDWKRVPNEPKLITT 9266
Qy      7162 ---TPTLQRPQEQATMRKFSLGGRGVAGVGYCTFAFGDAGMGLQGPMWARIANAV 7217
Db      9267 CECVFNLUKE-----LKYRF-----RVK-AV 9287
Qy      7218 SQSEEEQEEARAEQSEEQEAFARASPLQVSARVPVEVG-----RAPTRS 7264
Db      9288 NEAGESPDDTTEIPATDIOEE-----PEVFI---DIGAQDCLVCKAGSQIRIPAVI 9337
Qy      7265 SPEPTP---WEDIGQVSLVQ---INDLSGDAEADTISLDISEVDPAVNLNSLDYDI--K 7316
Db      9338 KGRPTPKSWEFDGKAKKAMKGDVHDIPEDAQLETAENSSVIIIPCKRSHTGKYSITAK 9397
Qy      7317 YLPPFEMIFRKVPKSAQPEPPSPMAEELAEFP-EPTWPWPGELG----- 7360
Db      9398 NKAGQKTANCRVKVMVDPGPPKDLKVSDITRGSCLSKWMDDDGCDRIKGYVIEKRTID 9457
Qy      7361 -----PHAGL-----EITESESDVALLAAAVGRKRWKSWSPSRLSFFHFGRLPL 7406
Db      9458 GKAWTKNPDGCGSTTFVVVPDLLSEQQYFVRVRAENRFGIGPPVETIQTATTARDP--IYPP 9515
Qy      7407 DEPAELGLRERVKASV-----EHSIRILKGRPEGLEKEG----- 7440
```

```
Db      9516 DPPIKLTIGLITKNTVHLSWPKPKNDGSGSPVTHYIVECLAWDPTGTGKEAWRCNQKRDVE 9575
Qy      7441 -----PBRKKPGPLASPRLSGLKSW-----DRAPTF-LRELSDETYYL 7476
Db      9576 ELQFTVEDLVGGEYEFVRKAVNAAGVSKPSATVGPCCQCPDMPDPSIDLKEFME--VEE 9633
Qy      7477 QGSYTLACQVSAQAQATWSKDGAPLESSSRVLISATLKNFQLL--TILVVVAE----D 7530
Db      9634 GTNVNIVAKIKGVFPPTLTWFKAPPKPDNKEPVLVDTHVNKLVVDDTCTLVIQSRSSD 9693
Qy      7531 LGVTTCVSNALGTVTGTVLRKAEPRSSSPCPDIG-----EYVADGVLLVWKPVESVGP 7585
Db      9694 TGLYITITAVNNLGTASKEMRLNVLGRPG-----PPVGPPIKFESVSADOMTSLWFFPKDDGG 9749
Qy      7586 ---VYIIVQ-CSLEGGSWTTIASDIFDCCYLTSLKSRGGTYTERTACVSKAGMG-PYSSP 7640
Db      9750 SKITNYVIEKREANRKTWVHVSSEPKECTYIPKLLGHEVYVFRIMAQNKGIGIEPLDSE 9809
Qy      7641 SE 7642
Db      9810 PE 9811

RESULT 2
I38346
C:Species: Homo sapiens (human)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C:Accession: I38346
R:Labelit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38346
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PID:g101
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match      8.0%; Score 3313; DB 2; Length 7962;
Best Local Similarity 21.5%; Pred. No. 4.4e-88;
Matches 1710; Conservative 1105; Mismatches 3287; Indels 1868; Gaps 306;

Qy      10 PRFTTRPKAFVSVGKDATLSCQIVGNPTPOVSVKEKQQPVVTAGARFLAQDGLYRLTI 69
Db      10 PAITPLQDVTVTSRGQPARFCQVSGTDL-KVSWYSKDKKIKPSRFRMTQFEDTYQLEI 68
Qy      70 LDLALGDSGVVCRARNAGEAPAAVGLQVDAEAACAEQ-----APHFLLRPT 117
Db      69 AEAPVEDEGTYTFVANNVAGQVSSSTANLSUAEPSILHERIEQEIEMEMKAAPIVKRIE 128
Qy      118 SIRVREGSEATFCRVGSGSPRPAVSWSKDGRRLGEPDGPVRVRVEELGEASALIRAAARPR 177
Db      129 PLEVALLHLAKFTCEIQSAPNVRQWFKAGRIEVESDKCSIRSKY--ISSLEILRTQVV 186
Qy      178 DGGTYEVRNENPLGAASAAALVVDSDAADTASRPGTSTAALLAHLORRRREARAGAPA 237
Db      187 DCGEYTCASNEYGSVSTATLV-----TEA 213
Qy      238 SPSTGTGRTCTVTE--GKHARLSCVYTGEPKPKVTWKKDGLQVTEGRRHVVVEYDAQENFV 295
Db      214 YPPTFLSRKSLTTFVGAAKAFICTVTGTVIETIWDKDGALSPSPNRI--SREKNKH 272
Qy      296 LKILFCQSRGRLTYCTASNLVG-QYTSVLLVVRREPAPVPPKKRLQDLEVRKESATFLC 354
Db      273 LELSNLTIQDRGVYVSCASNKFGADICQAEILIIDKP--HFIKELEPVQSAINKKVHLEC 330
Qy      355 EVPOP-STEAAWKEETRLWASAKYIEEGTERLTVRNVVSADDDAVYICETFE---GS 410
```

Db 331 QVDEBRKVTWTKDQKLPCKDKYKICPEDKIATLPLAKLKDSTGTVCTASNEAGSS 390
Qy 411 RTVAELAVQ--GNLRLKPRKTAVRVGDTA-MFCVELAVPVPVHWRNQBVEVAGGRAVA 467
Db 391 SCSATVTVREPPSFVKVDPSPVYMLPGESARLHCKLKGSPVQVTFWFKNNKELSESNTVR 450
Qy 468 ISAEGRTHLTISQCCLEDDVGQAFMA---GDCQTSRFRVCVAPRKPPLOPPVDPVKA 523
Db 451 MYFVNSEALLDITVKVEDSGSYCEAVNDVGDSCSTEIVIKEP--PSFIKTLSPADIV 508
Qy 524 RMESVILSWSPPPHGERPVITDGLVBEKKLGTWTWIRCHEABEWATPELTAVDVAEBEG 583
Db 509 R-----GTNALLQC-----EVSCTG 523
Qy 584 NQFRVSALNFGQSPYLEFPQGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDVLTVASAGE 643
Db 524 PFEI-----S 528
Qy 644 WFLDQALKASSVYBIHCDTRHRTLTIREVPASLHQAQKLVANGIESSIRMEVRAAPCL 703
Db 529 WFKDKKQIRSSKYLRFQKSLVCLUEIFSNAGDVEYECVAVNEVGKGCWMATH----- 583
Qy 704 TANKEPPAAAAAR--EYLABLHEBAQLLAEL-SDQAAAATWLDKGRTLSPCKPYEQASAGR 760
Db 584 LLKEPPTFKVKVDLIALGGQVTVLQAAVRGSEPISTWMMKQEVIREDDKIKMFSNGV 643
Qy 761 RVLLVRDVARDDAGLYECVSR---GGRIAY-OLSVQGLARFLHKDMAGSCVDVAVAGGPAQ 816
Db 644 AVLIIQVQISFGGKYTCCLAENAGSQTSGVGLIVKEPAKIER---AELIQVTAGDPAT 700
Qy 817 FECETS-EAHVHVHVKDGMELGHSERFLQEDVGTGRHLVAATVTRQDEGTY-----SCR 871
Db 701 LEYTVAGTPELKPWKYKDGRLPVLAG-KKRIISFKNNVAQLKPYSAELHDSGGQYTFEISNE 759
Qy 872 VGEDSVDFRLRVSEPKV--FAKEQLARRKLOAEAGASATLSCEVA-OAQTEVTWYKDGK 928
Db 760 VGSSECTTFTVLDRIAFFTK---PLRNVDSVNGTCRLDCKIAGSLPWRVSWFKDGK 816
Qy 929 KLSSSKVCMEATGCTRLVVOQAQADAGEYSCBA---GQRLSFHLDVKEPKVFP-- 982
Db 817 EIASDRIYAFVEGTASLEIIRVDMNDAGNFTCRATNSVSGSKDSSGALIVQEPSPFVTK 876
Qy 983 --AKDOVAHSEVOAEGANATLSCEVAQAAB---VMWYKDGKLSLKVHVEAKGR 1036
Db 877 PGSKDVLPGSAV-----CLKSTFGSGTPTLIRWFKNGKELVSGGSCYITKEALE 925
Qy 1037 RRLVVQAGKTDAGDYSCB---ARGQVSRFLHITPEPKMFAKQSVHNEVOAEGASA 1092
Db 926 SSLELYLVKTSDSGTYTCKVSNVAGVECSANLKVKEPATFVEK---LEPSQLLKKGDAT 982
Qy 1093 MLCSEV-AQAQTEVTWYKDGKLSLSSSKVGMVEKGCCTRLVLPOAGKADAGEYSCAGQ 1151
Db 983 QLACKVTGTPPIKITWIFANDREIKESSKHRMSFVESTAVLRITDVGIEDSGEYMCQAQNE 1042
Qy 1152 RVSEH---LHITPEKGVAKQSVHNEVOAEGATTAMLSCVA-OPQTEVTWYKDGKLS 1207
Db 1043 AGSDHCSIVIKESPYTKE---FKPIEVLEKEYDVMLLAEEVAGTPPEITWFKDNTILR 1099
Qy 1208 SSSKVRMEVKGCTRLRVVOQVKADAGEYSC---EAGQVSRFOLHITPEKAVPAKEQL 1263
Db 1100 SGRKYKTFIQDLHVLSQLILKFPVADAGEYQCVRTNEVGSSICSAVTLREPPSPFKK--- 1156
Qy 1264 VHNVRVTRAGASATLSCEVAQAQTEVTWYKDGKLSLSSSKVRIEAGCMRQLVVOQAQ 1323
Db 1157 IESTSSLRGGTAFOATLKGSLPITVTLWLDKSDDEITEDNIRMTENNVAASLYLSGIEVK 1216
Qy 1324 DAGEYTC---EAGQVSRFOLHITPEKAVPAKEQLAHRKVAQAGATLSCEVAQAQ 1379
Db 1217 HDGKYVQAKNDAGIQRCSALLSVKEP-ATTIEEAVS---IDVTQGDPTALQVKSFGTK- 1271
Qy 1380 EWT--WYKDGKLSLSSSKVRMEVAGCTRLRVVOQAQADAGEY-----SCEAGG 1426
Db 1272 EITAKWFKDQGBELTIGSKYKISVTDVTSILKIISTEKDSDGEYTFEVQNDVGRSSCKA-- 1329

Qy 1427 QRLSPSLDVAABPKVVPFAKEQPVHREVQAOAGASTTILSCVEAQAQ-TEVMWYKDGKLSLSS 1485
Db 1330 -RINV-LDLIIIPP-SFTKKL---KQWDSIKSGSFIDLECTIVAGSHPISIOWPMDQDEISAS 1383
Qy 1486 SKVRMEVAGCTRLRVVOQAQADAGEYSC---EAGQVSRFOLHITPEKAVPAKEQAS 1541
Db 1384 EKYKPSFHDNTAFLEISOLEGTDSTYTCSSATNKAGHNQCSGHLTVKEPP--YFVEKPS 1441
Qy 1542 REVQAEAGTSATLSCEV-AQAQTEVTWYKDGKLSLSSSKVRMEVAGCTRLRVVOQAQAD 1600
Db 1442 QDV--NPNTRVQLKALVGTAQMTIKWFKDNKELHSGAARSVMKDDTSTLSLEFAAKATD 1499
Qy 1601 AGEYSCK---AGDORLSFHLHVAABPKVVPFAKEQPAHREVQAOAGASATLSCEV-AQAQ 1655
Db 1500 SGTVICQLSNDVGTATSKATLTVKEPPQFIKKPSPV---LVLNRNQSTTFEQITGTPKI 1556
Qy 1656 EVTWYKDGKLSLSSSKVRMEVAGCTRLRVVOQAQADAGEYSC---AGQVSRFOLH 1711
Db 1557 RVSWYLDGNEITAQKHGISFIDGLATFQISGARVENSQVYCEARNAGTASCIELKV 1616
Qy 1712 ABLEPQISERPCREPLVVKHEDIILTATLATPSAATVTLKOGVEIRRRKRHETASQ 1771
Db 1617 KEPPTFIRE---LKPVEVVKYSDVELECEVTGTPPEVTVLKNRREIRSSKKYTLTDRV 1672
Qy 1772 DTHILTVHGAQVLSAIYSCRVGABGQFPVQVEEVAAK---FCRLLEPVPVCGELGTVT 1827
Db 1673 SVFNLIHTKCPSDTGEYCIVSNEGGSCSCT-RVALKEPPSPFIKKIENTTTTVLKSSAT 1731
Qy 1828 LACEL--SPACAEVVMRCGNTOPVGRKFMVAEGP-----VRSILTVLGLRAED---AG 1876
Db 1732 FQSTVAGSPPIIS-ITWL-----KDDQILDEDDNVYISFVDSVATLQIRSVONGHSG 1781
Qy 1877 EYVCESRD---DHTSAQLTVSVPRVVKFMSGLSTVVAEEGGEATFQCWSPS-DVAVVM 1931
Db 1782 RYTQAKNESGVERCYAFLLVQEP--AQIVEKASVDVTEKDPMTLECVVAGTPELVKVM 1839
Qy 1932 FRDQALLOPSKPAISQSGASHLTSIDLVEDAGQITVEAB---GASS-SNALRV--RE 1985
Db 1840 LKDGKQIIVPSRYFMSFENNVSFRIQSMVKQDSQGYTFKVENDFGSSCDAYLVLQDN 1899
Qy 1986 APVLFKKLEPQTEERSVTVLELTPWP-ELRWTNRNATALAPGKVEIHAEGARHL 2044
Db 1900 IPPSFTKLLTKMDKVLGSSIIHMEKVSGLSIPISQWFKD-----GKEI---STAKVRL 1950
Qy 2045 VLNHVGFADRGFCGCTPDDTKOAKLTVMERQVRLVRGLQAVAREQGTATMEVQLSHAD 2104
Db 1951 VCH-----ERSVSLVNNLEL-----EDTAN----- 1971
Qy 2105 VDGSWTRDGLRPOQGPCHLAVRGPMTLTLTSLRPEDSGLMVFKAEGVHTSARLVVTEL 2164
Db 1972 -----YTCKVSNVAGDDA-----CSGILTVKEP 1994
Qy 2165 PVSPSRP-LQDVV--TTEKEKVTLECELSRPNVDVRLKDGVELBAGTKMTAAAGACRS 2221
Db 1995 PSFLVKPGRQAIIPDSTVEFKAILK---GTPPFKIKWPKDDVELVSGPKCFI 2051
Qy 2222 LTIYRCEPADQGVVYCDADHAQSSASVKVQGTVTTLIYRRVLAEDAGBIQFVAENASRA 2281
Db 2052 LNLISVDASKTGQYTC-----VTNDVG-----SDSCTT 2080
Qy 2282 QLRVKELPVTLLVRPLURDKIAMKRGVLECOVSRASQVR--WFKGSELQBPVKYELVS 2339
Db 2081 MLLVTEPPKFKVCKLEASKIVKAGDSRSLCKIA-GSPBIRVWVFRNEHELPAKDYRMT- 2138
Qy 2340 DGLYRKLIISDVHAEDEDTYTCADGVKTSQAQFFVEEQSIITVRGLQDVWMEPAPAFE 2399
Db 2139 -----FIDSVAVTOMNNLSTEDSG-----FI 2160
Qy 2400 CETSIPIVRPKLLGKTLVLAGGVNGLQEGTVHRLMLRRTCTMTGPHVFTVKGSRSS 2459
Db 2161 CEANQPA-----GST-----SCST----- 2174

Db	2526	KPKDVTALENAVVSFELUSGSHDTPVVRWFPHKNVELQSDQKYKMSQRKVHKLMLHNISP	2585
Qy	2711	EDAGLYTCHVGSEETRARVRVHDHLHVGITKELKTMWELEGESCFECVLSHESASDPAMW	2770
Db	2586	ADAGEYTAFUQLECKAKLFWETH--ITKTMKSIETPETKTASFQCEVSHFNV--PSVW	2641
Qy	2771	TVGKTVGSSSRFOATROGRKYIILVREAPSDAGEVVSVRGLTSKASLIVRERPAALI	2830
Db	2642	LKNGVEIEMSEKFKIVVQGLHQLNMNTSSDSAEYTFVCGNDRVSATLTV--KPLIT	2699
Qy	2831	KPLEDQWVAPGEDVELCELSRAGTPVHWLKDRAIKRSQKYDVVCGTWMVLVIRASL	2890
Db	2700	SMLEDINAEEDKDTITFEVTVNYEGISYKWLKNGVEIKSTDKQIRTKLTHLSLRNVHF	2759
Qy	2891	KDAGEYTCVEASKSTASHVEEKANCPTBELTNLOVEEKGTAFTVCTKTHPAATVTRK	2950
Db	2760	GDAEYSFVAGKAASSATLYVEARHIEPRKHIDKIVVEKKRALFECIESEPVOVQWMK	2819
Qy	2951	GLLELRASGKHQPQEGJTLRLTTSALEKADSDTYTCDIGQAQSRALLVQGRRVHIIED	3010
Db	2820	DQELQIGDRMKIQREKYVHRLIIPSTQMSDAGQTVVAGNTSSANLIVEGRDVRIRSI	2879
Qy	3011	LEDVDVQBGSSATFRCRISPANYEPVHFLDKTPHANELNEIDAQPGGY-----HVL	3064
Db	2880	RKEIQVIERQRAEIEFEVNEDDIEP-QWYKD-----GIEINFHYEERYSYVVERRIHMS	2933
Qy	3065	LRQALAKDSGTIYFEGDQORASALRVT--EKPSVFSRELTDATITGEDTLVLCETSTC	3122
Db	2934	IFETYSDBGYTFVAGNRNSVLYYNAPPEPQI--QELQPTVESGKPARFCALISGK	2992
Qy	3123	DIP-MCWTKGCKTLRGSAQCQLSHGHRQAQLLITCATLQDSGRYKCEA---GGACSSSI	3177
Db	2993	POPKVSMYKDQQLSPGFKCKFLHDAQEYTLILLITFPEDSAVYTCCKNDYGVAITTSAS	3052
Qy	3178	VRVHARPVRFQF-----ALKDLEVLGGAATLRCVLSSVAAAPVKMVCYGNVLR	3225
Db	3053	LSVRIPEVSPLEBPVPPYPAVIVPLRDVAVTSEQSARFQCRVTTGTLKYSWYSKOREIK	3112
Qy	3226	PGDKYSLRQEGAMLELVNRLRPODSRY---SCSPGDOTTSATLTVTALPAQFTGKLR	3281
Db	3113	PSRFRMTQFEDTYQLEIAEAYPEDEGTYTVASNSVGQVTTSTAILKLEAPEKIMYBKJE	3172
Qy	3282	NK-----EATEGATATLRCELSKTAPV--EMWKGSETLRDGDYCLR	3321
Db	3173	EEIEMEVKVAPILARRLEPLEVAVNHVAKFTCEVETTPNVKFWYKAGREIYDGKYSIR	3232
Qy	3322	QDGAMCELOIRGLAWDAAEYSVCGGBERTSAS----LTIRPMPAHFIBLRHQESIEGA	3377
Db	3233	SSNYLSTLEIPRPQVQDCEYSKASQHGSSVSTAFLTVTE--PPRFIKKLDSSRLVKQH	3291
Qy	3378	TAT-LRCEL--SKAAPVWRKGRESLRDGRHSLRQDGAVCELOICGLAVADAGEYSC--	3432
Db	3292	DSTRECKVGSGSPIKVTWYKGETEIIHPSEKYSMSFVDSVAVLEMHNLVSDESDGYSCEA	3351
Qy	3433	--VCGBERTSATLTKVALPAKFTFEGLRNEEAVEGATAMLCELSKVP--VEVRKGPENL	3488
Db	3352	QNPAGSASTSISLKVAKPPA--FTKKPHPVQTLKSDVHLECELOQTFFPQISWYKOKREI	3410
Qy	3489	RDGDYIILROBGTGRCLOICGLAMADAGEYLCVC---GOERTSATLITRALPARFIEDV	3544
Db	3411	RSSKKYKMSBENYLASIHILNVDTADVGEYHCKAVNDVGSDCIGSVTLRA--PPTFVKKL	3469
Qy	3545	KNQERAGATAVLOCELNSAAPVE--WRKG--SETLRDGDYSLRQDGTCKELQIRGLAMA	3601
Db	3470	SDVTVVVGETIELOQAAVEGAQPISVLWLOKGEIIRESENLIWISYENVASLJGNAEPT	3529
Qy	3602	DTGYSVCVGQERTSAMLTVRALPIKFTFEGLRNEEAEGATAVLRCELSQMAVPEWVKGH	3661
Db	3530	NAGXYIC-----	3536
Qy	3662	ETLRDGRHSLRQDARCELOIRGLVAEDAGEYLCMOGKERTSAMLTVRAMPSKFIETGLR	3721

Query Match 6.0%; Score 2489; DB 2; Length 6642;
Best Local Similarity 20.8%; Pred. NO. 2.5e-64;
Matches 1520; Conservative 972; Mismatches 2737; Indels 2064

Db 2284 E--SVKPSAIVTGKPMNPVWYLNKKLIQSBEVKVYVHETGKTSIRIQKPLMBHN--- 2338
Qy 2052 ADRGFFGCETP-----DDKTQA-KLTVEM--ROVRLVRGLQAVEARE-OQT 2093
Db 2339 ---GTRIVEAENVSGKVQATQALKVDKKEVFKFTNMDDRQVK-----EGEDVKFT 2387
Qy 2094 ATMEVOLSHADVGDGWRDGLRFQOGPTCHLAVRGMHTLTLGSLRPEDSGMLVPKABG- 2152
Db 2388 ANVE---GYPEPSVAMTLNGEVPVSKHPNITVTDKGEHTIEISAVTPEQAGELSCATNP 2444
Qy 2153 VHTSARLV-----VTELPVFSRPLQDVVTEKEKVTLCELS--RPNVDVRLMKDGV 2204
Db 2445 VGSKRDRVOLAVKKVGDAP-TFAKNLEDRLITGELTMDAKNLVTKPKKITWLKDGVE 2503
Qy 2205 LRA-GKTMIAAQAQACRSITIVRCBFADQGVVVCDAHD-----AQSASVKKVQGRVTYLIY 2259
Db 2504 ITSDGHYKIVEEDGSLKLSILOTKLEDKGRITIKAESEFGVAECASLGV-----VK 2556
Qy 2260 RVLVADAGEIOFVAENAESRAQLRVKELPVLTVRLPRDKIAMEKHRGVLECOV---SRA 2316
Db 2557 GRPMAPKA---FQSDIA-----PINLT-----EGDTLECKLLJITGDP 2590
Qy 2317 SAQVWFSGSOBLQPGPKVELV-SDGLYRKLIISDVHAEDEDTYTC---DAGDVKTSAQ 2371
Db 2591 TPFVRYIQTQVLCATEDTEISANGVY-TMKIHGVTADMTGKIKCVAYNKAGEVSTEG- 2648
Qy 2372 FVFEBSQSIIVRGLQDVVMEPAPAFWE---CETSI-----PSVRPPK 2411
Db 2649 -----PLKVVAIPVFEFETSLCDATCREGDTLKLRAVLLGBEPVW---S 2690
Qy 2412 WLLGKTVLOAGNVGLEBGTVHRLMLRRTCSMTGPPVHFTV-----GKRSARLIVGDI 2467
Db 2691 WTVNGKLEESNIKIHSEKGTYYTIIKOITCDYSGVVCEAINEYGHATSEATL----- 2745
Qy 2468 PVVLRPRLEP-----KTG-RELQSVVLSCDFRPAPKAVOMYKDDTPLSPSEKFK 2515
Db 2746 -LVLPGRGPPDLEWLSNVARTGTVKHVVFTGDPKP---SLTWYNNKEILNSDLYT 2801
Qy 2516 MSLEQMAELRILRMP-----ADAGVYRQOA-----GSAHSSTEVTVARE 2557
Db 2802 IVTDKKTSTLTINSFPDVHVGEIICKAENDAGEVSCVTANMITYTSDMFSESESAQAEE 2861
Qy 2558 -----VTVTCP-----LODAEATERGWAFSPCLSHEDDEE---VE 2589
Db 2862 FVGDDLTDESUREEMHRTPTVMAKPFITIKDKTKAKKHGSAVFECVW---PDTKGVCCK 2919
Qy 2590 WSLNGMPLYNDSFHIEISHKGRHT-----LVLSKISORADAG---IVRASSLVK 2634
Db 2920 WLKDGKEI-----ELIARIVQRTGPEGHITQELVLDNVTPEBAGKYTCIVENTAGD 2973
Qy 2635 STSARLEV-----RVKPVVFLKALDDLGAERGTIALQCEV-SDPEAHVVMRKDGV- 2684
Db 2974 TCEATLTVIESLEKSEKKAPEFIVALQDKTKTSEKVVLECKVIGBPKPVSMLHDVVS 3033
Qy 2685 -QLGPSDKVDFLHTA-----GTRGLVVHDVSPEDAGLYCHV-----GSSETARVRVH 2732
Db 3034 REKNPSEKTTIQESITVESVEGVERVTTITSELSHOGKYTCIAENTEGTSKTEAFLTVQ 3093
Qy 2733 DLHVGITKRLKTMVELEGESCFEVLBSHASDPAMMTVGKKT-----VGSSSRF--- 2783
Db 3094 GEAPVFTKELQNKELSIGEKVLVSCVKGSPQPHVDYFSPSETTKVETKITISSRIAIEH 3153
Qy 2784 -QATRQGRKYI-LVVREAPSDAGEVFSVGLTSKASLIVREPAALIKPLEDOWVAPG 2841
Db 3154 DQTNHWMVISOITKEDIVSYKAIATNSIGTATSTSKITTKVEAPVFEQGLKTKTSVKEK 3213
Qy 2842 EDVELRCLSRAGTVPVHMLKDKRAIRKSQYDVV---CEGTWAMLVIRGASLKDAGEYTC 2899
Db 3214 BEIKMEVKGSGAPDVEWFKDDKPVSEGDNGHEMKKNPETGVFTLLVWQAAATTDAGKYTAK 3273
Qy 2900 VEASKSTASLHVEKAN-----CFTBEL--TNLQVEKGTAVFTCKTEHPAAVTVWRKG 2951
Db 3274 ASNPAGTAESSAEVTOLEKPTFVRELVTTEVKINETATLSVTYKGV-PDPSVEMLKD 3332

Qy 2952 LLELRASGHQPSQ-EG-LTLRLTISALEKADSDTYTC-----DIQQAOSRAQLLVQGRRV 3005
Db 3333 GQPQTDSSHVIKVEGSGSVSITI KOARLEDSGYACRAINPAGEAKTEANFAVKNLV 3392
Qy 3006 --HIILEDVDDVOEGGSATPRCRISPNYEPVWFILDKTPLHANELNEIDAQPG-GVHV 3062
Db 3393 PPBEVEKLSPLLEVKEKSTLTSVKVVGTPPEPSWFWKDDTPISIDNVHVIQKQTAVSFS 3452
Qy 3063 LTLQALKDSGTIYF-----EAGDQRASAAALRVTEK--PSVFSRELTDAITIGEDITL 3115
Db 3453 LTIINDARQDVG-IYSCRARNEAGEALTANFGIIRDSIPPEFTQKRLPLEVREOETL 3511
Qy 3116 VCETSTCIDP-MCWTKDQKTLGRSARCQLSH-----EGHRAQLLITGATLODSGRYKC 3167
Db 3512 KVTYIGTVPVNVWFKDKPI---NIDNSHIFAKDESGHHT-LTIKQARGEDVGVTYC 3566
Qy 3168 ----EAGACSSSIVRVHAR---PVRFOEALKOLEVLEGGAAATLRCVLSVAAP-VKWCY 3219
Db 3567 KATNEAGEAKTTANNAVQEEIEAPL-FVQGLKPYEVEQGPAPLAVVRVEGKPEPEVKWFK 3625
Qy 3220 GNNVLRPGDKYSLRQEG-AMLELVVRNLRPQDSGRYSC---SFGDQTTSATL----- 3267
Db 3626 DGVPIADNQHVIIEKKGENGSHLLVIKDTNNADPGKYTCQANTKAGKDETVGEUKIPIKYS 3685
Qy 3268 ---TVTALPAQFTIGKLRNKEATEGATATLRCELSKTA--PVEWRKGSSETLRDGDYCLR 3321
Db 3686 FEKQTAEEVPLFTIEPLKETFAVEGDTVVLECKVKNKESHQIKPFKNDQPVIEIQHMJLE 3745
Qy 3322 --QDGMCELOIRGLAMWDAABYSC-----VCGEBRTSASLTI----- 3357
Db 3746 VLEDGNI-KLTIONAKKEDYGAVRCEAVNVAGKANTNADLKIQPAAKVEEHVTDESQLE 3804
Qy 3358 -----RPMPAHFIGRLRHOBSEIEGATATLRCELSKAAP---VSWRK 3395
Db 3805 EIGQFETVGTASKTDTGAP-EFVELLRSCVTVEKQQAILCKV-KGEPRPKIKWTK 3862
Qy 3396 -GRESLRDGRHSLRQDGAVCLEIQICGLAVADAGEYSCVCGEERTSA-----TLTVKAL 3448
Db 3863 EGKEVMSARVRAHBKDDGTLLTFDQVNTQADAGEYRCEAENEYGSATWETPIIVTLEGA 3922
Qy 3449 P-----AKTEGLRNEEAVEGATAMLCWELS-KVAP-VIEWRKGPENLRDGDYILR--Q 3498
Db 3923 PKIDGEAPDFLPQPKPAVTVGTAVLEKIGISKPKPSVKYKNGEELKPEPSDRVKIENLD 3982
Qy 3499 EGTRCELOICGLAMADAGEYLCVCGOE---RTSATLITIRALPAR---FIEDVKNOEAR 3550
Db 3983 DGTQ-RLTVTNAKLDDMDEYRCEASNEFGVDVSDVTLTVKE-PAQVAPGPFKEISATQVK 4040
Qy 3551 EGATAVLQCELSAAP-VIEWRKSETLRDGR--YSLRQDGTKCELOIRGLAMADTGEY- 3606
Db 4041 ETETAKPECKVSGTKPDVKWFKDGTPLKEDKRVHFESTDDGTQ-RLVIEDSKTDQGNYR 4099
Qy 3607 ---SCVCQGBRTSAMLTVRALP---IKFTEGLRNEEAEGATAVLRCEL-SKMAPVEMWK 3659
Db 4100 IEVSNDAGVANSKVLTV--VPSETLKIKGLDQVNVVTQGTKILLSVEVEGKPKTVKWKY 4157
Qy 3660 GHETLRDGRHSLRQ-DGARCELOIRGLVAEDAGEYLCMCGKE---RTSAMLTVR--- 3710
Db 4158 GTETVTSSTQTKIVQVTESEYKLEIESAEMSDTGAYRVVLSTDSFSVESATVTVTKAAE 4217
Qy 3711 --AMPSPKIFEGLRNEEAEGDTATLMCEL-SKAPVVIEWRKGHETLRDGRHSLRQDGSRC 3767
Db 4218 KISLPS-FKKGLADQSVPGKTPLVLEVEIEGKPKDVKYKNGEIKDQKVEDL--GNGKY 4274
Qy 3768 ELOIRGLAVVDAGEYSCV---CGOERTSATLTVRALPARFIEDVKNOEAREGATAVLQC 3823
Db 4275 RLTIPOFEKQDVGSEYSVTAAEAGEIESKAKVNVSAKP-EIVSGLVPTTVTKQGETATFN 4333
Qy 3824 EL-SKAAPVIEWRKSETL-----RGGDRYSLRQDGTTRCELOIHLGLSVADTGEYSCV-- 3873
Db 4334 KVGKPVKGVKWKYKNGKEIPDAKTKNDGSGYS-----LEIPNAQVEDAADYKVVS 4384

Db 6038 -----EVNGPEE-----CKPR----- 6049
QY 5945 YVFRNMKLSIDLNDQVEGDDRAPEWQEREDSVRKQLLOARTALIKSSWKEICGIQQ 6004
Db 6050 -TRGLYNS-----IHE----- 6061
QY 6005 RLALPVRPPDFEELADCTABELGTVKLACRVGTGPKPVIWYKDGKAV---QVDPHHI 6061
Db 6062 -----GNVEMIVCATGTPPTVKWYKDGQEIIVGDGDKRV 6098
QY 6062 LIEDPDGSCALLDSLTGVDSCQYMCFAASAGNCSTGLKILV-----QV 6106
Db 6099 IFTDERGIHLLVINVASPDDEGEYSLEATNKLGSATGSLNIIRPRIADADERGMPF 6158
QY 6107 PRFVNVKVRASPFVEGEDAQFTCTTEGAPYPOIRWYKDGALLTGNKFTGLSEPRSLLV 6166
Db 6159 PGFVRQLKNKVFHNMPTIFDCLVVGHPAPEVFNHNGKLVPGRIK-IQSCGGGSHA 6217
QY 6167 LVIRAASKEDLGLYBCELVNRLGSARASAEIRIQSPMLQAOBQCHREQLVAAVEDTTLER 6226
Db 6218 LIILDTTLEDAGEYVATAKNHSGSASSAVLDVTVPFL-----DSIKFN 6261
QY 6227 ADOEYTSVLKRLG-----PKAPGSTGLTGGP-----PCPRGAP----- 6262
Db 6262 GEIDVTPYLTEYGFKKLNTASLPTPPDRGPFKEVTHGYLTLFWIPTGRAPPYQVTV 6321
QY 6263 -----ALOETGSQPPVT-----GTSE-APAVPP----- 6284
Db 6322 VIEIRELPEKWSLLEYNIPEPVCKVRNLELGSQFVRRAENIYIGIDPSPASPRLM 6381
QY 6285 RVQPILLHE-----GPEQEPBAIARAQB-----W-----TYQAIDHTA----- 6354
Db 6382 APPQPVFORRTNKVPLLDPAEAKALDMRYSOYACAPWFSQVVEKRYCAENDTLTIVL 6441
QY 6313 RMEG-----AAMPAGTCELLWDVSHVVRETTQRT-----YTYQAIDHTA----- 6354
Db 6442 NVSGFPDPDKWFRG-----WDI-----DTSSETSKCVYTYGSGSETTLAIGTFSKEN 6490
QY 6355 -----RPPSMQVETIEDVQAQGGTAQ-----FEAITEGDQP 6386
Db 6491 VQYQCFKNDYGDQAQNMVDLATPFPFIQLVN-----KTFSSAQPMRMDVRVDGEFP 6546
QY 6387 SVTWYKDSVLVDSRSLQQEGT-TYSLVLRHVASKDAGVYTCIAQNTGGQVLCKABELL 6445
Db 6547 ELKMKWKEWPVSESRIRKFDQGPLCSLIINDPMWRDSGIYSCVAVNDAGQATTSCTVT 6606
QY 6446 VLGDNEDPSEKQSHR-----RKLHSFYEVKEE 6473
Db 6607 VEAEGYNDVELPRRVRTIESRRVRELYEISEK 6639

RESULT 5
T43290
hemictinin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text_change 09-Jul-2004
A:Accession: T43290; T20993; T24734
R:Voegel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A:Description: Hemictinin is required for hemidesmosome mediated cell adhesion and germ-
A:Reference number: 222396
A:Accession: T43290
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5198 <WIL>
A:Cross-references: UNIPROT:076518; EMBL:AF074901; PIDN:AAC26792.1
R:Suiston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: 219355
A:Accession: T20993
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <WIL>

A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <WIL>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: him-4; F15G9.4b
A:Map position: X
Query Match 5.0%; Score 2048; DB 2; Length 5198;
Best Local Similarity 21.4%; Pred. No. 1.1e-51;
Matches 1187; Conservative 773; Mismatches 2284; Indels 1292; Gaps 254;

QY 286 VYEDAQ-----ENFVLKILFKOS-----DRG-----LYTCTASNLVGQTY-- 321
Db 136 VFTDARSKDHLDEVLTITQEKOSVYFVMTGDCGNRTHGPFRTYEKIAAASFGVPHL 195

QY 322 -----SSVLVVVRBA-----VPPFKRLQDLEVR-----EKES 349
Db 196 EKSDSVTLEVYRVHVAQKKVHLMYEAREGGTVSRNIPVDKHLSELTISLGGDKSDN 255

QY 350 ATFLCEVPQPSAEAMFKEETRLWASAKYGIIEBGTERRLTVRNVSADDDAVYICETPEG 409
Db 256 LDIVLRDPEGRT-----VDKRLYS-----KEGGTDLKNKILRLKDPSPGVTWNTN 303

QY 410 SRTVAELAVQGNLLRKLPRKTAVRVGTAMFCVELAVPGVPHVLRNQEVEVAGRVAIS 469
Db 304 SRLKHTIRVGHGAVDPKYGFASRPLDR-----IELAR-PPV---LNQDTYLLINMTGLI 355

QY 470 AEGTRHTLTISQCCLEBDVGQVAFMAGDCQTSRPFVSAAPRKPPLQPPVDPVVKARMSSV 529
Db 356 PPGT-----VGEIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPV----- 394

QY 530 ILSWSPPHGEPVITIDGYLVEKKKLGTYTWIRCHEAEVATPELTADVADAEQNFQFRV 589
Db 395 -----PPKGLFFRVQGY-----DEDNYEFMR 416

QY 590 SALNSFCQSPYLEFPPTVHLAPKLAVRTPILKAVQAVEGGEVTFPSVDLTVASAGB--WFLD 647
Db 417 IAPTAI-----GSIIVGGPRAFMSPIH---QEFVGRDLNLSCTVESASYTYVWKT 465

QY 648 GOALKASSVYEHCDRTRHTLTIREVPASLHGA---OLKFFVANGIESS1--RMEVRAAPG 702
Db 466 GEDIIGGLFYHNTDTS--VWTIPEL--SLKDAGEYECRVISNNGNYSVKTRVETRESP- 520

QY 703 LTANKPPAAAREVLEARLHEEAQLLAELSQAIAA---VTWLKDGRTISPGPKYEVQASAG 759
Db 521 -----PEIFGVRNVSVPLGEAFL--HCSTRSAGEVIRWTRYGATVFNQNTERNPTNG 573

QY 760 RRVLLVRDVARDDAGLYECVSR--GGRIAYQLSVQGL-----ARFLHDKMAGSCVDAYAGG 813
Db 574 --TLKIHVVTRADAGVYECVARNAGMSTRKMLDIMEPPSVKVTPODVVFNRMREGV--- 628

QY 814 PAQPECET-SEAHVHVHVKYKGMELGHSGRFLQEDVGTGRRHLVAAATVTRDQEGTYSRV 872
Db 629 --NLSCEAMGDPKEVHVHVKYKGMELGHSGRFLQEDVGTGRRHLVAAATVTRDQEGTYSRV 872

QY 873 -----GEDSVDFRLRVSEPKVFAKEQLARRKLQABAGASATLSCEVAQAQ--TEVTWY 924
Db 684 MSQAGQARDTDLML-ATPPKV-----EIIQNMVWGRGDRVSEFCKTIRGKPKIRWF 737

QY 925 KDGKLLSSSKVCMKMEATGCTRLVWQOAGQADAGEYSC-----BAGGRLSFHLDV-KEPK 979
Db 738 KNGKOLIKPDYIKINEG---QLHIMGAKDEADAGAYSCVGENMAGKDVQVANLSVGRVPT 794

Db	2728	V--RSQVEHAGTYTCEAQNNGKARKDFLVRVTAPPH	Db	3739	NAGLYTCIALNRAGEASLEFKVEILSPVPIODISRNDVOP-QVAVNQPTIMRCVGTGHPFP
Qy	2934	VFTCKTEH--PAATVTW-----RKGLLELRASGHQPOBGLTLRLTISALEKADSDT	Qy	3919	TVVMKSGLOLQANGRRPRLQGGTAEILAVLOLQREDTGEYTC
Db	2779	LITCNAESSVPLSSVWHAHDESQNGVI-----TSKYAANEK-----TLNVTNIQLDDRGF	Db	3798	SIKMLKNGKEVTDDENIRIVEQGVLOILRTD--SDHAGKWSCVAENDAGVKELEWMLDV
Qy	2985	YTCIDGQASRAQLLVQRRVHIIEBLEDVVOQE-----GSSATFRCCRISPANYEPVHW	Qy	3975	TAAVRFURELQHOEVB-----GGTAHLCCELS-RAGASVEWRKGSLOLPFCAYQMVO
Db	2831	YVC-----TAVNEAGITKFKFLIVETPVPFLOQKLPIILGKRLTLDC--SATGTPPPTI	Db	3856	FTPPVSVKS-----DNPIKALGETITLFCNASGNPYPQLKWAGGSLIFDS-----P
Qy	3039	FLDKTPLHANELNEIDAQGGVHVLTQLALKDSGTIYFE--AGQORASAALRVTKEPS	Qy	4029	DGA-----AELLVRGVEQEDAGDYTCO-----TGHTQSMASLSVVRPCK-----FKT
Db	2886	LFMKDGRKLNESDEVI--IGSTLVINDNPKQKEGRYTCTAENKAGSEKMMVVELLPK	Db	3904	DGARISLKGARLDIPLHKKKTDVGYTCOALNAGTSEASVVDVLVP-PEINRGIDMSP
Qy	3097	VFSRELTDTATIEGBDLTVCE-TSTCDIPMCWTKD-GKTLRGSARCOLSHEGHRAQLLI	Qy	4073	RLQLEQBTGDIARLCCOLSDAESGAVVQWLKEGVEL-HAGPKYEMRSQAGATRELLIHQL
Db	2945	L-SKEWINVEVOAGDPLTLECPIDETSGVHIITWSQFGDKGQDWMRAQSSD--KSKLYI	Db	3963	RLPAQOSLT-----LQC-LAQGKVPQWRWTLNGTALTHTSTPGITVASDSIF--IQINN
Qy	3155	TGATLQDSGRYKC-----BAGGACSSIVRVHARVPRFOEALKDLEVLGGAAATLRCVLSS	Qy	4132	EAKDTGEYAC-----VTGQKTAASLRVTEPEVTIVRG---LVDAEVTADEBVEFSCVSR
Db	3002	MQATPEDADSYCIAVNDAGGAEEVQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRTEG	Db	4015	SLSDKGVYTCVAENVAGSDNLMYNDVVQAPVISNGGTQVIEGELAVIE-----CLVEG
Qy	3211	VAAP-VKWCY-GNNVLR-PGDKYSLROEGAMLELVNMLRPDQSGRYSCSGDQTTSA--	Qy	4185	AGATGVQWCLQGLPLOSNEVTEVAVRDRGRIHTLRKLGVTPEAGTIVSFLGNHASSAQL-
Db	3062	IPPPBISWFLDGKPILEMPGVTY---KQGD-LSLRIDNIKPNQEGRYTCTVAENKAGRAEQ	Db	4070	YPAPQVSWLRNGRNVETGVQGVRYVTDRM--LTIIEARSLDSGLYLCSATNEAGSAQA
Qy	3266	-TLVTALPAOPFIGKLRKEATEGATATLRCEL--SKTAPVEWRKGSSETLRDGDY--C	Qy	4244	---TVRAPEVT-----ILEPLQDVQLSEGQDASFQCRLSRASQGEARWALGGVPLQANE
Db	3118	DYVIEISPPRVMASEWRVVEGQTTIRCEVGNPEPVMN-----LKDGEYTSDL	Db	4128	YTLVILVSPKLIITSPGVLT-----SSGSKFSLPCAVRGYPDPPIISWTLNGNDIKGE
Qy	3320	LRQDAMCELOIRGLAMVDAEYSCV---CGEERTSASLTIRPMPAHFIRGLRHOESI-	Qy	4295	MNDITVEQGLHL-----LTLHKVLTLEDAG--TVSFHVGTCSSEAOQLKYAKNTVVVRG
Db	3172	LQFSKLSVHLRETTLADGGTYTCTAATKAGESQTTTDEVLVPP-----RIEDEERV	Db	4182	NGHTIGADGTLHIEKABERHLIYECTAKNDAGADTLEFPVQT-----IVAPKISTSG
Qy	3375	---EGATATLRCELS-KAAP-VIEW-RKGRESLRDGRHSLRDGAVCBLOICGLAVADAG	Qy	4346	LENVEALEGGEALPECOLSQPEVAHAHTWLLDDPVRTSENAEVVFFENGLRHLLKNLR
Db	3227	OQKEGNTVMVHCQVTRPVYVYWRNGKE-----IEQFPNV--LHIRNATRADEG	Db	4234	NRYINGSEGTETVIKCEI-ESESEFSWSKNGVPLPFSNN--LIFSEDYKLIKILSTRLS
Qy	3429	EYSCVCGEERTSAT---LTKALPAKTEGLRNEEAVEGATAMLWCELS--KVAPVEWR	Qy	4406	PQD--SCRVTFLAGDMVTSAPLTVRGWRLEITLEPLKNAAVRAGAQAARFTCTLSBAVPGE
Db	3276	KYSCIASNEAGTAVADFLIDVFTKPT-FETHTETTNIVEGESAKIECKIDGHPKPTISWL	Db	4291	DOGEVSCCTAANKAGNAQTKTNLV-GVAPKIMERPRQTVVHKGDQVTLWCEAS-GVPQPA
Qy	3483	KG--PENLRDGDYLRQEGTRCELOICGLAMADAGEYLCVC---GOERTSATLIRAL	Qy	4464	ASWYINGAAVOPDDSDMTVTADGSHQALLRASAOPHHAGEVTFACRDAVASAR---LTV
Db	3335	KGRPFNM--DNIIILSPRGD--TLMILKAORFDGGLYTCVATNSYGDSEQDFKVVVYTK	Db	4349	ITWYKDNELLTNTGVDETAIT--KKKSIVFSSISPSOAGVYTCKAENWVASTEEDIDLIV
Qy	3537	PARFIEDVNO--EAREGATAVLOCEL--NSAAPVEWRKGSSETLRDGRYSLRQDGTKE	Qy	4520	LGLPD-PEDAEEVVAHSSHTVTLSSWAAPMSDGGGLCGYRVEVKEGATQWRLCHELVPG
Db	3390	P--YIDETIDQTPKAVAGEIILKCPVLGNPTPTVTKRGDDAVPNDSRHTIVNN--YD	Db	4407	MIPPEVVPERNVSTNPRQTVFLS-----
Qy	3593	LQIRGLAMADTGEYSCV---CGOERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCE	Qy	4579	PECVVDGLAPGETYRFRVAAGVPVGAGEPVHLPQTVRLAEPPKVPQPSAPESRQVAAG
Db	3445	LKINSVTTEDAGQYSCIAVNEAGNUTTHYAAEVIKPTFVRKGNLYEVIENDITLMDCG	Db	4431	--CNATGI-----PEPV-----
Qy	3649	LSK--MAPVWVKGHETLRDGRHSLRDQGARCELOIRGLVAEDAGEYLCMGKERTSA-	Qy	4639	EDVLSLELVABAGEVITWHKEME-RIQPGGRFEVVSQGRQOMLVIKGTABEQGEYHGL
Db	3505	VTSRPLPSISWFRGDKPVLYDRYSISPDGS--HITINKAKLSDGKVKICRASNEAGTSD	Db	4441	-----ISWMRDSNTAIONNEKYQILG---TTLAIRNVLPPDDDGPHYC-I
Qy	3706	--MLTVRAMPSKFTIEG--LRNEEATEGDTATLWCELS--KAAPVEWRK-GHETLRDGRH	Qy	4698	AQGSICFAAATFQVALSPASVDEAPQPSLPPPEAAQEGDLHLLWEALARKRMSREPTLDS
Db	3563	IDLILKILVPPKIDKSNIGNPLAIVARTIVLECPISGIPQPDVITWTKNGMDINMTDSRV	Db	4481	AKSDAGOKIATRIKLVNKPDRPAP-----IW-----
Qy	3759	SLRDGSRCELQIRGLAVVDAGEYSCV---GOERTSATLTVRALPARFIEDVKNQEAR	Qy	4758	ISELPEEDGRSRLPQBAEEVAPDLSEGYSTADELARTGDADLSHTSDDDSRAGTSLV
Db	3623	ILAQNE--TGIEINVQVTDQGRYTCTATNRRGGKASHDFSLDVLSPPEFDIHGTPTIKR	Db	4508	-----VECDE-----KGPCKT
Qy	3815	EGATAVLOCELSKA-----APVEWRKGSSETLRGG--DRYSLRQDGTRELCQIHGLSVA	Qy	4818	TYLKKAGRPGTSPKASKVGAPAPSVKPOQOQEEPLAARVPLGLDLTKDLGDPNDKAAV
Db	3681	EGDITTLTCTPIKLAEDIAQVMDVSWTKDSRALDGLDNDVDSDDGRK--LTISQASLE	Db	4520	EYMDIRG-----DTPDDNPQL-----LPWKVDVSDSLNGSIA
Qy	3866	DTGEYSCVCGOERTSATLTVRA---PQPVF---REPLQSLQABEGSTATLOCELS-EPTA	Qy	4878	KIQAAFGYKVKRKEMKQOEGFMESHHTFGDTEAQVGDALRLCEVVASKADVBARWLKQAVE
			Db	4552	--YRCMPGPRSSRTVLLHAAPQFIVPKPNTTAAIGAIVELRCSAAGPPHPTITWAKOGL

1618 HVAEPKVVFAKEQ--PAHREVQAEAGASATLSCVAQAQTEVTVYKDKGLSS--SSKVR 1673
1400 -----IVDAKNKFVIALKDTVEIKDDVTLMCQTKDTKPGIWRNGKQISSMPPGGKPE 1453
1674 VEAVGCTRRLVVQQAQADAGAYSCEAGQORLSFRLHVAELE--PQISERPCR-----RE 1726
1454 TQSRNGHTLTKIGKIEEMADYVEIDQAGLRSCNVTVLEAKRPILNWKPKKIEAKAGE 1513
1727 PLUVKEHEDIILTATLATPSAATVTLKDGVBIRSKRH--ETASQGDTHLTVHGAQVL 1784
1514 PCVVVPFQIKGTRR--GDPKAQI-----LKNKGPIDBEMRKLVEIILKDDVAEIVFKNPQLA 1569
1785 DSAIYSCRVG-----AEGODF-----PVQVEEV 1807
1570 DTGKWALBELGNSAGTALAPPELVFKDKPKPPKPLETKNVTAEGLDLVNGTDPDDEGAPV 1629
1808 AAKFCRLLE-----PVGELGGTVTLACELSPACAEVVMR-----CQNTQPRVGRKFPQ 1855
1630 KAYIITEMQGRSGNNAKVGETKGTDFKVKDLKEH--GEYKFRVKALNECGLSDPLTGE--S 1686
1856 MVAEPVRSRLTVLRAEDAGYVCESDRDHTSAQLTVSVPRVKFMGSLTVVAEBEGE 1915
1687 VLAKNP-----YGVPGPKPNMDAIDVDKDHCT-----LAWEPPEEDGG 1724
1916 ATFQCVVSPSDVAVVWFRDGLALQSEK-----PAISQSGASHSLTISDLVLEDAQOI--T 1969
1725 API-----TGYIIERREKSEKMDHQVQGTQKPCCBELTKKKVVEDKEYLYR 1769
1970 VEAEGASSAALRVREAPVLFK--KLEPO-----TVEBRSSVTELEVLT--RPWPEL 2018
1770 KVAVNKAQGDPCDGHGKPKMAKAKASPEFTGGGKDLRLKVGETIKVDVPISGPELPEC 1829
2019 RWRNATALAPQKNVEIHAEGARHLVLHNVGADRGFF-----GCETPPDDKTOAK 2069
1830 LMVNGKPLKAVGRVKMSSERKHKIMKIENAVRADSGKFTITLKNSSGSCD-----STAT 1884
2070 LTVEHROVRLVRGLQAVEAREOGTATME-----VOLSHADVGSWTRDG 2113
1885 VTVVGRPTPKGPDIAVDCADG--ATLSMNPDDDDGPDPLTGYIVEAQMDNKGKYEYG 1943
2114 LRFOQGPCHLAVRPMHTLTLSGLRPEDSGLMVFKAEGVHTSARLVTVELPVSFRLQ 2173
1944 -----KVDPTNTTLKNGLR--NKGNYKFRVKAANNGE-----SEPLS 1980
2174 -DWVTEKEVLTCELSRP-----NVDRVW--LKDQ-----VELRAGKTMAI 2213
1981 ADQYTOIKDPWDEPGKGRPEITDFDADRIDIAWEPHKGCGAPIEEYVEVRDPDK-- 2038
2214 AAQACRSLTIVRCEBADOQVVCDAHDAQSSASVKVQGRVTLIYRVRVLAEDAGEIQPV 2273
2039 -----EWEK-----VKRPVPTNAGISGLKEGKEYQFRVRAVNKAQPGQ----- 2076
2274 AENAESRAQL--RVKELPVTVLVRPLDKIAWEXHGRVLEQCVSRASA-----Q 2319
2077 -PSESEKOLAKPKFIPAWL-----KHDNLKSTVKAQATVRWEVKIGGEPIPE 2124
2320 VRWFGSOELQPGKYELVSD-----GLYRKLILSDVHAEBE----- 2356
2125 VKWFKGNQOLENG--IQLTIDFRKNEHTILCIPSAWRSVVGVEYR--LTVKNSHGADEKAN 2181
2357 ----DITYCDAGDVKTSQAFFVEBOSITIVRGLQDVTVNRP-----APAWFEC 2400
2182 LTVLDRPKPNPGLVSDVF-----EDNLNLSWPPDDGGEPTIYVEVEKLDATGRWVPC 2238
2401 -----ETSIPTSV-- 2407
2239 AKVKDTKAHIDGLKQGTQYQFRVKAANNKESDALSTDKDTKAKNPYDEPGKGTGPDVVD 2298
2408 -----RPPKWLKGLKTVLQ-----AGNVGLE-----OEGTVH 2434
2299 WDADRVSLWEPPKSGDGAIPITQYVIEKKHGRDQWECGKVGSGQDTNAEIILGLKEGBEY 2358

2435 RLMLRRTCTWTGTPVHFTVGKSSRSARLVVSDIPVVLTRPLEPKTGRELQSVVL-----S 2489
2359 QFRVAV--NKAGP-----GEASDPSRKVAVK-----PRNLKPWIDREAMKITITIKVND 2406
2490 CDF-----RPAPKAVOMYKDDTPLSPSEKFKMSLEGQMAELRILRLMPADAGVYRCQA-- 2542
2407 VEFDPVVRGEPKPK--EWIFNEKPPV--DQIRIESEDYKTRFVLRGATRGAGLVTLTAT 2464
2543 ---GSAHSTEVTVPEAREVTVTGPIQDABATEEGHNASFCELSHEDEEVEWSL-----NGM 2595
2465 NASGSDKHSVEVILGKPSPLPLEVSNVVE-----DRADLEWKVPEDDGA 2512
2596 PLYNDSPEHISHK-----GRHHTLVLSIQRADAGIVRASLSKVSTSRLEVRVK 2645
2513 FI---DHYEIEKMDLATORWPCGRSET-----TKTVPNLQPGHEYKFRVR 2556
2646 PVVFLKALDLSAE-----ERGTALQCEVSDPEAHVVMRKDGVLQ----- 2687
2557 AVNKGESDPLTNTTNTAILAKNPYEPVPGVKDPPEL--VMDKDHVDLAWNAPDDGGAPIEAF 2615
2688 -----PSD-----KYDFLHTAGTRGLVHVDVSPEDAGLYT 2717
2616 VIEKDKNGRWEALVWPGDQKTATVPLNKEGEYQFRISARNKAGTGDSPDPSD----- 2670
2718 CHVSEETRARVRVHDLHVGITKRLKTMEVLEGESCFEVLSHESASDPANMTVGGKTV 2777
2671 -RVVAKPRNLAPRIH-----REDLSDTTVKVGATLKFIVHIDGEPAD--VTWSFNGKI 2722
2778 GSS-----SRF-----QATROGRKYILVREAAAPSADAGEVVSFVRGLTSK----- 2817
2723 GESKAQIENEPVSRFALPKALRQSGKYTITATNINGTSDVTINIKVSKETPKPGPIE 2782
2818 ASLIVRERPAIILPLEDO-----WV----- 2838
2783 VTDVEDRATLDWKPPEDDGGEPPIEFYIEKMNKTDGIVWPCGRSGDTHFTVDSLNGKDH 2842
2839 -----APGEDVELCEL-----SRAGTP-----VHW--LKORK 2864
2843 YKFRVKAVNSGSPDPLETETDILAKNPFDRPRPEPTDMSDHDVLDKMDPLDSDGG 2902
2865 A-----IRKSKQYD-----VVCETMAMLVIRGASLKDAGEYTCV-----E 2901
2903 APIEYQIEKTKYGRWEPAITVPGQTTATV--PDLTPNEEYFRVVAVNKGSPDPSD 2960
2902 ASKTASLHVBEKANCFTTELTNLQVBEKGTAVFTCKTE--HPAATVTVR--KGLLELRASG 2959
2961 ASKAVIAKPRNLKPHIDRALKNLTIKAGQSISFSDVPVSGEPAPTIVTWHWPNREIRNGG 3020
2960 KHQPSQEGTLRLTISALEKADSDTYTCDIGOAQSRALLVQGRVHIE-----D 3010
3021 RVKLDNPEYQSKLVVKQMERGDSGFTTIKAVNANGEDEATV---KINVIDKPTSPNGPLD 3077
3011 LEDVD-----VQEGSSAT----- 3023
3078 VSDVGHGHTLVNWRAPDDDDGGPIENYVIEKYDVTASGRVWPAKVAGDKTTAVVDGLIPG 3137
3024 ---FRCRISPANYE---PVHWF---LDKTPLE-----HANELNEID---AQF 3057
3138 HEYKPRVAANAEGSDPLETFTGTTAKDPDPKPKTNAPEITDMDKOHVDLEWKKPPAND 3197
3058 GG-----YHVLTLRQLA---LKDSGTIYF-----EAG---D 3082
3198 GGAPIEEYVEMKDBFSPFNDVAHVAPAGQTNATVGNLKEGSKYBFRIRAKNKAAGLDPS 3257
3083 QRASAAALVTEKPSVFSR--ELTDATITBEGEDLTLCVETSTCDIP--MCWTCKGKTLRGSAR 3140
3258 DSASAVAKARVPPVIDRNSIQEIKVKAQDPSLNIPIVSGEPTPIITWTFTGTPVESDDR 3317
3141 QQLSHGHEHRAQLLIITGATLQDSGRY--KCEAGGACSSSIVRVA-----RPRVFORA 3190
3318 MKLNNEDGKTFHVKRALRSDTGTIYIIKAENENGTDTAEVKTVDLHPSPRGLDVTNI 3377
3191 LKDLVL-----EGGAATLRCLVSSV--AAPVKV-----CYGNN----- 3222

Db 786 ---GGEKKPM-TEAERQSLFPFGKVEKWDIPLPEKTVOQVDKICEWKCTYSRPAKI 840
Qy 1382 TWYKDGKLSSSS---KYRMEAVGCTRRLLVVOQAQADTGEYSCEAGGQRLSFLSDVAP 1438
Db 841 RMYKDKRIFSGGLKYKIVIEKNVCT--LIINPEVDDTKYKTCIANGVPTHQAULTVLEP 898
Qy 1439 KVPFAKEOPVHREVQAQAGASTTLCSEVAQAQTEVMYKDGKLSFSSK---VRMEAVG- 1494
Db 899 PMKYSNPLPNTQBIYTKQAVLTCKVNTPRAPLVYRGSAIQOEGDPRFTIIEKDAVGR 958
Qy 1495 CTRRLVVOQAQADAGEYSCEAGSQRLS-FHLHVAEPKAVFAKEQPASREVOAEAGTSAT 1553
Db 959 CT--LTIKEVEDDAEWARTITQDVFSKVQVYVEEPHTFVPMKSQKYNESDL---AT 1013
Qy 1554 LSCEVAQAQTEVTWYKDGKLS--SSSKVRMEAVGCTRRLLVVOQAQADAGEYSCKAGDOR 1612
Db 1014 LETDVNDKDAEVMVWHDGKRIDIGVKFVSSNRKRLLIINGARIEDHGEYKCTTKDDR 1073
Qy 1613 LSFHLHVAEPKVPFAKEQ--PAHREVOAEAGASATLSCEVAQAQTEVTWYKDGKLS-- 1668
Db 1074 TMAQL-----IVDAKNKPIVALKDTEVIEKDDVTLMCQTKDKTPGIFRNGKQISSMP 1127
Qy 1669 SSKVRMEAVGCTRRLLVVOQAQADAGEYSCEAGGQRLSFLHVALE--POISERPCR-- 1724
Db 1128 GGFETQSRNGHTLTKIGIEMNEADVTEIDQAGLRGSCNVTVLEAKRPIILNWKPKKIE 1187
Qy 1725 ---REPLVVKEHEDIILTATLTPSAATVWLKOGVEIRRSKRH--ETASQGDTHLTAVH 1779
Db 1188 AKAGEPCVVVPFIKIGTRR-GDPKAI--LKNGKPIDEEMKRLVIEIKDDVAEIVFK 1243
Qy 1780 GAQVLDSAIYSCRVG-----AEGQDF-----PV 1802
Db 1244 NPQLADTKWALELNSAGTALAPPELVFKPKPKPGKPLETKNTVTAEGLDLVWGTPDPD 1303
Qy 1803 QVEEVAAPFCRLLE-----PVCGLGTVTLACELSPACABVWVR-----CGNTQPRV 1850
Db 1304 EGAPKAYIIEMQEGSRGNWAKVGTGTDFVKOLKEH-GEYKFRVKNALNECGLSDFLT 1362
Qy 1851 GKRQFQWAEVGRSLTVLGLRAEDAGEYVCESRDDHTSAQLTVSVPRVVKPMGSLTVVA 1910
Db 1363 GE--SVLAKNP-----YGVPGKPKNMDAIDVDKDCHT-----LAWBPP 1398
Qy 1911 BEGGEATQCVCVSPSDVAVVFRDCAILOPSEK-----PAISQSGASHLSTISDLVLEDA 1965
Db 1399 BEDGAPTI-----TGYYIERREKSEKWHQVGTQPCCELTKDKKVEDK 1443
Qy 1966 QOI-TVEAEGASSAALVREAPVLFK-KLEPQ-----TVEERSSVTLEVELT-R 2013
Db 1444 EYLVRKAVNKAAGPCDCHGKPIKNKAKKASPEFTGGGKDLRLKVGETIKYDVPISGE 1503
Qy 2014 PPELRLWTRNATALAPGRKNVEIHAEGRHRLVHLNVGFADRFF-----GCETPDD 2064
Db 1504 PLPECLVWVGKPLKAVGRVMSRGRKHIMKIENAVRADSGKFTILKNSSGSCD--- 1559
Qy 2065 KTOAKLTVMRQVRLVRGLQAVEAREOQTATME-----VOLSHADVDGS 2108
Db 1560 -STATVTVGRTPPKPLDIAVCAADG-ATLSMNPDPDDGDPDLTGIVIEAQDMDNKGK 1617
Qy 2109 WTRDGLRFQOGTCHLAVRGPMTLTLGLRPEDSGSLMVFAEGVHTSARLVVTELPSVF 2168
Db 1618 YIEVG-----KVDNPTTTLKUNGLR--NKGNYKFRKVAKNNEGE----- 1654
Qy 2169 SRPLQ-DVVTTEKVBTECELSRP-----NVDVRW---LKDG-----VELRAG 2208
Db 1655 SEPLSADQYTIQKDPWEPGPGREBITDADRIDIAWEPHKDGGAPIEBEYIIVEVRDP 1714
Qy 2209 KTMATAAGACRSLSITYCEPADQGVYVCDADHAQSSAVKVGQRTYTLIVRVLAEADG 2268
Db 1715 DTK-----EWKE-----VKRVPDTNASISGLKEGKEYQFRVAVNKAAGPG 1754
Qy 2269 EIQFVAENAEAPQAL-RVKELPVLTVRPLRDKIAEKGRLVLECOVSRASA----- 2318

Db 1755 Q-----PSESEKQOLAKPKFIPAWL-----KHDNLKSIITVKAGATVRWEVKIGG 1798
Qy 2319 ---QVRWFKSQELQPGPKYELVSD-----GLYRKLIISDVHAD 2355
Db 1799 EPIPEVKWFKNGQLENG--IQLTIDTRKNEHTILCIPSAMRSDVGEYR-LTVKNSHGAD 1855
Qy 2356 E-----DIYTCADGDKTSAQFFVEEQSITIVRGLQDVTVMEP-----AP 2395
Db 1856 BEKANLTVLDPRSPKNGPLEVSDVF---EDNLNLSWKPPDDDDGGGPIEYVEYEKLDATG 1912
Qy 2396 AWFEC-----RPPKWLKGTVLQ-----AGNNGVLE-----O 2429
Db 1913 RWVPCAKVKDKAHLIDGLKKGQTYQFRVKAVNKEGASDALSTDKDTKAKPNPDEPKTGT 1972
Qy 2405 PSV-----RPPKWLKGTVLQ-----AGNNGVLE-----O 2429
Db 1973 PDVVDWADRAVSLSEWEPKSGGAPITQYVIEKKGKGRDQWQCEKVGSGDOTNABILGK 2032
Qy 2430 EGTVHRLMLRRTCSMTGPHFTVCKSSRSLRVVSDIPVLTTRPLEPKTGRELOSVL- 2488
Db 2033 EGEEYQFRKAV--NKAGP-----GEASDPKRVKAK-----PRNLKPIDREAMKTTITI 2080
Qy 2489 ---SCDF-----RPAPKAVQWYKDDTPLSPSEKFKMSLEGQMAELRILRLMPADAGVY 2538
Db 2081 KVGNDVEFDVPVRGEPPEPK-EMFNEKPVD-DQKIRIESEDYKTRFVLRGATRKHAGLY 2138
Qy 2539 RQCA---GSAHSTEVTVAREVTVTGPLQDAEATEBEGWASFSCELSHEDBEVMSL- 2592
Db 2139 TLTATNAGSGDKHSVEVIVLGPSPLEVSNYE-----DRADLEWKVPE 2186
Qy 2593 --NGMPLYNDSEFHEISHK-----GRHRTLVLKSIQADAGIVRASSILKVSTARS 2640
Db 2187 DDGGAPI---DHYEIEKMDLATGRWVPCGRSET-----TKTTPVNLQPGHBY 2230
Qy 2641 EVRVKPVFLKALDLSAE---ERGTALQCEVSDPEAHVVRKDGVOQLG----- 2687
Db 2231 KFRVAVNKEGSDPLTNTAILAKNPYEVFGKVDKPEL-VDMKDKHVDLAWNAPDDGA 2289
Qy 2688 ---PSD-----KYDFLHTAGTRGLVHVHDVSPED 2712
Db 2290 PIEAFVIEKDKNGRWEALVVPGDQKTATVPLNKEGEYQFRISARNKAGTGDPSPD 2349
Qy 2713 AGLTYCHVSGSETRARVVDLHVIGITKRLKTMVELEGESCFECVLSHESADAMTV 2772
Db 2350 ---RVAKPRNLAPRIH-----REDLSDTTVKVGTATLKFIHIDGEPAPD-VTVSP 2396
Qy 2773 GGTGVGSS---SRF---QATROGRKYILVVREAAAPSADAGEVFSVRGLTSK- 2817
Db 2397 NGKGIGESKAOIENEPIYSRFPALPKALRKQSGKYITATNINGTDSVTINIKVKSPTKP 2456
Qy 2818 ---ASLIVRERPAIILPLEDO-----WV----- 2838
Db 2457 KGPIEVTDPEDRATLDWKPPDDGGEPIEFVEIEKMTKDGIVWPCRSRGTHTFTVDSL 2516
Qy 2839 ---APGEDVELRCEL-----SRAGTP-----VHM--- 2859
Db 2517 NKGDIKFRKAVNKEGSDPLETETDILAKNPFDRPRGRPEPTDWDSDHDLKWDPP 2576
Qy 2860 LKDRKA-----IRKSQYD---VVCETGMAMLVIRGASLKDAGEVTCV- 2900
Db 2577 LSDGAPITEEYQIEKRTYKGRWEPAITVPGGQTATV--PDLTPNEEYFRVAVNKGCP 2634
Qy 2901 ---EASKSTASLHVEEKANCFTBELTNLQVEEKGTAVFTCKTE-HPAATVTWR-KGLLE 2954
Db 2635 SDPSASKAVIAKPRNLKPHIDRDALKNLTIKAGQISFDVPVSGEPAPTVMHWPDRNE 2694
Qy 2955 LRASGHOPSEGLTLRLTISALEKADSDTTCIDGOAQSRAQLLVQRRVHIE----- 3009
Db 2695 IRNGRVLKNDPEYQSKLVVQMERGDSGTFTIKAVNANGEDATV---KINVIDKPTSP 2751
Qy 3010 ---DLEVDV-----VOEGSSAT----- 3023
Db 2752 NGPLDVSVDHGDHVTLNWRAPDDGGGPIENYIEKYDTASGRWVPAKAVAGDKTTAVVD 2811

Db 4636 RPKDNGGDAITVYVVEKRTPGCDWVT--VGHVPGVTTLRNLDANTVPEFRVRAENQYG 4693
Qy 4898 PMFSHTFGDTEAQVGDALRL-----ECVVASKADVARWLK---DG 4935
Db 4694 -----VGEPLTDDAIKAKNPDPFGAPGQPEAVETSEEAITLQWTRPTSDG 4740
Qy 4936 -----VE-----LTDGRHHHDQGDGTCSSLLIAGLDRADAGCYTC 4971
Db 4741 GAPIQYVIERKEVSGSTWTKAAGNILDTHRVTLGTPKTYEYFRVAAYNAAGQGEYS- 4799
Qy 4972 QVSNKFGQVTHSACVVSSESESSGELDDAFRAARELHRLFR--TKSPAESVDEE 5030
Db 4800 -----VNSVPIADANAPTRKINMGMLTRDILAYAGERAKILVPPAASPA---PK 4846
Qy 5031 LPLSADEGPAEPEPADQVREDEHFCIRPEALTEARQAQVTRFQEMFATLIGIVEIKL 5090
Db 4847 VTFSGENKISPTDRPVKVEYSDFLATLTIEKSELTDGGLYFVELENSQGSASIRLKV 4906
Qy 5091 VEQ--GPRVEMCISKETAP-----VVPPEPLSLTSDAPV--FLTELQNEQVQDG 5140
Db 4907 VDKPASQHIRV-----EDIAPDCCTLYWMPPS-----SDGSPITNVIKELDLRHS DG 4956
Qy 5141 -----YVVSFDC-----VVTGQP- 5153
Db 4957 KWEKVSSFVRNLNYTVGGLIKDNRFRVRARBTQYGVSEPCBELADVVAKYQFVNPQPE 5016
Qy 5154 MPSV-----RWFKDGKLLJEEDDHYMINBDQ 5178
Db 5017 APTVRDKDSTWAELEWDPDPRGSKIIGVQVQYRDTSSGRWINAKMDLSEQCHARVTGLR 5076
Qy 5179 QGQHOLIITAVVPADMGVIRCLAENSMGV-----STKAELR-----VDLTS-- 5220
Db 5077 QNGE-----FBRILAKNAAGFKSPSPSERCOLKSRFGPGPIHVGAKSIG 5124
Qy 5221 -----TDYDTAADATESSEYFSAQYLSRREQEGTESITDE----- 5256
Db 5125 RNHCITWMAPLEDGGSKITGVNIREYGSTLWTVASDY--NREPEFTVDKLRFNDEYE 5183
Qy 5257 -----GOLPOVBEELDLQVAPGTRLAKFOLKVKGYP 5289
Db 5184 FRVVAINAAGKGIPLSPGPIKIQESGSRPQIVVKPED-TAQPNYRRAVFTCEAVGRPE 5242
Qy 5290 PRLYFKDGQPIATSAHRTMKKILHTLEIISVTRDSGQVAAVISNAMGAAYSSARLL 5349
Db 5243 PTARLNRNGRELPESSRYRFEASDGVYKFTIKEVMDIDAGEYTVESVNPYSGSDTANLV 5302
Qy 5350 VRGDPPEEKPASDVHEQOLVPRMLERFTPKVKKGSSTITFSVKVEGRPVPTVHVL---R 5406
Db 5303 VOAPPVIEK-----DVNTILP-----SGDLVRLKIYPSG-TAPPRHSLVLNR 5344
Qy 5407 BEAERGVLWIGDTPGYTVASSAQHSLVLLDVGRHQOQTYTTCIASNAAGQALCSALHV 5466
Db 5345 EE-----IDMDHPTIRIVEFDDHILITIPALSVREAGRYEYTVSNDGSEATGTFWLVN 5397
Qy 5467 SGLPKVBOEKVKEALISTFLOGTTOAISAOGLTASPADLGGQKBEPLAAKEALGHLS 5526
Db 5398 TGLPEAPO-----GPH 5409
Qy 5527 LAECTEERFLOKLTQITEMVSAKITQAKLVPGGDSDEDSKTPSPSRHGBSRSSSIQ 5586
Db 5410 ISNIGPSTATLSWRPPVTD-----GGSKITSYV 5438
Qy 5587 ESSSESDGDARGEIDFYVVTADYLPGLAEQDAILTREGQYVEVLDAAPLRLMLVTRKP 5646
Db 5439 EKROLSD-----EMVTVTSNVKMDNYIVT--- 5463
Qy 5647 TKSSPSROGWSPAYLDRRLKLSBPWGAAEAPFEPGEAVSEDEYKARLSSVITQLLSSEQ 5706
Db 5464 -----GLFE-----NHEYEFVSA----- 5477
Qy 5707 AFVEBELQFQSHLQHLERCHPVTAVAGQKAVIFRNVDRDIGRPHSSFLQELQCCDQDD 5766

Db 5478 -----QENEGIG----- 5484
Qy 5767 VAMCFIKNAAEQYLEFLVGRVQAESVWVSTAIQEFYKKAABEALLAGDPSPQPPPIQ 5826
Db 5485 -----APLVS-----EHPILARLPDPTPSPL- 5506
Qy 5827 HYLEQPERVQRYQALLKELIRNKARNRQNCALLQEAQAVVUSALPQRAENKLHVLSMENY 5886
Db 5507 -----NLBIVGGDYVTLWSQR----- 5524
Qy 5887 PCTLALGEPTRQGHFIVWEGAPGARMKGNHRRHVLFRNHLVICKPRDSRTDTSV 5946
Db 5525 -----PLSDG-----GGRL-----RGYIV--EKQBEHEDEWFC 5551
Qy 5947 FRNMMLSSIDILNQVEGDDRAFEVWQREDSVRKYLQARTAIKSSWVKEICIGQORL 6006
Db 5552 NQNPSPNNYVNPILDG-----RKY--RYRVFVANDAGLSLAELODTL 5594
Qy 6007 --ALPVRPPPEEBELADCTAELGETVKLACRVGTGPKPVISWYKDGKAVQVDPHILLIE 6064
Db 5595 FOAGSGEGPKIVSPLSDNEEVGRCVTFECBISGSPREYRMPFKGCKELVDTSYKTLIN 5654
Qy 6065 DPDGSICALILDSLTGVDSGOYMCFAASAAGNCSTLKGILVO-----VPPRPVNVKVRASP 6118
Db 5655 KGDQK-VLIINDLTSDDADEYTCRATNSGSTRSTRANIRIKTPRVFPKTHGGEAQ- 5712
Qy 6119 FVEGEDAQTCTIEGAPYQIIRWYKDGALLTTGNKFTQLSEPRSGLLVLVIRAAKEDLG 6178
Db 5713 --KGETIELKIPYKAYPQGEARWTKDGEKIENNSKFSITDDK--PATLIRISNAREYDG 5768
Qy 6179 LYECELVNRLGASARASABELRIQSPMLQAOEQCHREQLVAANVEDTTLERADQEVTSVLKRL 6238
Db 5769 EYRVVENSVDGSDGTNVTV----- 5789
Qy 6239 LGPRAPGSTGLTGPGPCPRGAPALQETGQPPVTGTSEAPVPRVP--OPLLEHGP- 6295
Db 5790 -----ADPVE-PRPPIIENIDEAVI 5810
Qy 6296 -EQPEAI---ARAQEWTVPIR-MEGAAMPAGAGTGELLWDVHSHVVVRETTRTYTYQAI 6350
Db 5811 LSWKPPALDGGSLVNTYTIKREAGGSM-----SPCAKSRITYTTIE 5853
Qy 6351 -THTARPSMQVTTEDVOAQT---GGTAQFAIIEGDPQSPVTVTKDSVOLVDSTRLSQ 6406
Db 5854 GLRAGKQYEFRIIAENKHGQSKPCPTA--PVLIPGDERKRRRGYD---VDEQCKIVR 5906
Qy 6407 QEGTYSVLVRHVASKDAVVYTCLAQNTGGQVLCKAELLVLGGDNEPDSEKSHRRKLHS 6466
Db 5907 KGTIVSS-----NYDNYVFDIWKQYY-----PQPEIKHDH---VLD 5940
Qy 6467 FYEVEKEIGRGVFGVKKRVQHKGNKILCAAKFIPILRSRTRAQAVR-ERDILAAALSHPLVT 6525
Db 5941 HYDIEHELGTGAFGVVHRVTERATGNNFAAKFVMTPHESDKETVRKEIQTMSVLHRPTLV 6000
Qy 6526 GLLQOPETRKTLILILELCSSEELLDRLY-RKVVTVEAEVKYIQQULVEGLHYLHSHGV 6584
Db 6001 NLHDAFEDDNEVMVITYEFMSGGELFEKVADEHNKMSDEAVEYMRQVCKGLCHMHENNV 6060
Qy 6585 HLDIKPSNLMVHPAREDIKICDFGAQNTIPAEQLQFQSGSPSPSPSPSPSPSPSPSPSP 6644
Db 6061 HLDLKPENIMTTTKRSNELKILDFGLTAHLDPKQSVKVTGTGAEPAAEVAGKPVGYT 6120
Qy 6645 DIWANGVISYLSLTSSPAGESDRATLLNLEGRVSWSSPMAAH--LSEDAKOFIKATL 6702
Db 6121 DMWSGVLSYILLSCSLSPFGGENDDETLRNVKS--CDNMDDSAFSGISEDGKDFIRKLL 6178
Qy 6703 QRAFOARPSAAQCLSHPWFLK-SMPABEAEHINTKQLK 6739
Db 6179 LADENRTMIHQALEHPMLTFCGNAPGRDQSPSSRYTK 6216

protein unc-22 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: A88852
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999
A:Accession: A88852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <STO>
A:Cross-references: UNIPROT:Q23550; GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN000000000
C:Genetics:
A:Gene: unc-22
A:Map position: 4
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;
Query Match 3.9%; Score 1619; DB 2; Length 6831;
Best Local Similarity 18.6%; Pred. No. 3.7e-39;
Matches 1442; Conservative 933; Mismatches 2680; Indels 2712; Gaps 325;
QY 120 RVREGSEATFRCKVGGSPRPAYWSKDGRRLLGEPDGPVRV-VEELGEA---SALRIRAR 175
DB 7 KCKQG-DLLMECHLEADPOPTIAQHSGNLL-EPGSRVVQTLPLGGSLYKATLVKEPN 64
QY 176 PRDGGTYEVRANPIGASAAALVVDSDAATASPGTSTAALLAHQRREARAGA 235
DB 65 AGDGAYKCTARNQLGESNANLNFAGAGGDEAKSRGPFV-----GK 108
QY 236 PASPSTGTTRCTVTGEGHARLSVVTGEPKPTWKKDG-OLVTEGRRHVVED-AOEN 293
DB 109 PRILPDG-----GALIVCKVKSASTVAKMKDGVPLSMGLYHALFSLDGLQDT 160
QY 294 FV-LKILFCQSDRLGYTCTASNLVGYTSVSVLVVREPAVPFKRLQDLVREKESAT 351
DB 161 YLCQLEIRGSSDAGQYRCNIRNDQGETNANLNFEEP-----DPSEQRKRST 212
QY 352 FLCEVQPSQTEAWFKEETRLWASAKYIE-EGETRLTRVNSADDDAVYICETPGS 410
DB 213 ---ASPRSSRPGGSRP-----SSPKSMKSRGTPKR-----TLKPRGS 250
QY 411 RTVAELAVQGNLLRKLPRKTAVRVGDGTAMFCVELAVPVGPVHMLNQBVEWAGGRVAISA 470
DB 251 PS-----KKLRSSTSPV-----NEEV-----SQS 270
QY 471 EGTRHTLTISQCCLLEDVGVAFMAGDCQSTSTRFCVSPARKPPLOPPVVPVKARMESSVI 530
DB 271 ESRRSSRT-----DKMEVDQVSG-----ASKRK----- 294
QY 531 LSWSPPHGERPVTIDGVLVEKKLGTYTWRICHEAENVAHPETLVADVAEGNPPQFVS 590
DB 295 -DGLPPPG-----DEKKUR-----AGSP----- 312
QY 591 ALNSFGQSPYLEFPFGTVHAPKLAVRTPLKAVQAVEGGEVTFPSVLDLTVASAGEWFLDQQA 650
DB 313 ---STRKSP-----SRKSASPTSRKSSAG-----AASGTTGASASA 348
QY 651 LKASVYIEHCDTRHTTITREVPASLHGAQLKFVANGIESIRMEVRAAPGLTANKPPA 710
DB 349 TSATSGGSSDASDRKYT-----RPP----- 370
QY 711 AAAREVLARLHEEAQLLAELSQQAAAVTWLKDGRTLSPGPKVEQASAGRVLVLDVAR 770
DB 371 -----IVLEASR 377
QY 771 DDAGLYECVSRGGRIAYQLSVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVH 830
DB 378 -----SQTGRI-----GGSVVLEV-----QMOCHSSTI-----IEW 403
QY 831 YKGMELHSGERFLOEDVGTFRHLVAATVTRQDEGTYSRV-----GEDSVDFRLRVS 884

DB 404 YRDGTLVRNSEE-YSQSPNGSIAKLQVNKLTEKSGLYKCHAKCDYGGQSSAMVYKIQS 462
QY 885 EPKVVFQKEQL-ARRKLOABAGASATLSCEVAQAQTEVTWYKDKKLSKSSKVCMEATGC 943
DB 463 DVEBELMKHRKDADEYQKEQKQTL-----QAEIKRVARRSKSKSKS----- 507
QY 944 TRRLVVOQAQADAGEYSCAGGQRLSPHLVDKPEKVVFAKDQVAH---SEVOABAGANA 1000
DB 508 -----PAPAKKSTTS-ESGRQBAS-----EVEHKSSSVRPDPDDES 544
QY 1001 TISCEVAQAQAEVVMYKDGKLSLKVHVBEAKGCRRLVVQQAQAGTCTDAGDYSCEARGQR 1060
DB 545 QLD-EIP-----SSGLTIPEE-----RRRELLGQVGSDD----- 573
QY 1061 VSFRLHTEPKMFAKEQSVNEVQABAGASAMLSCEVAQAQTEVTWYKDG----- 1111
DB 574 -----EVSEISBELPSFAGGPKRRKTDK 597
QY 1112 -KLS-----SSSKVGMVKGCTRLVLVLPQAKADAGEYSCAGGQSVFHLHTEPKGVF 1166
DB 598 PKKUSIAPVSTNKSSDDEPSTPRR-----RSSIDMRRESVOEILEKTSPLVPSGASGA 652
QY 1167 AKESVHNEVQAEAGTTAMLSCVA-QPQTEVTWYKDKKLSKSSSKVR---MEVKGCTRRLL 1223
DB 653 PKIVEVPENVTVVENETAILTKVSGSPAPTRFRWFKSGREVISGGRFKHITDGKEHTVAL 712
QY 1224 VVOQVGKADAGEYSCA-----GGQVSFQLHITPEKAVFAKEQLVHNEVTRTEAGSATLS 1279
DB 713 ALLKCRSQDEGPFYTLTIENVHGTDSADVKLLVTSNGLDFRAMLKHRH---SQAG----- 764
QY 1280 CEVAQAQTEVTWYKDKKLSKSSSKVRTEAAGCMRQLVVQQAQADAGEYTCAGGQRLSF 1339
DB 765 -----FQKDG-----EGGAGGG-----GGEKKPM 784
QY 1340 HLDVSEPKAVFAKEQ-----LAHRKVQABAGAIATLSCEVAQAQTEVTWYKDKKLS 1392
DB 785 -TEAERQSLPPGKKVEKWDIPLPEKTVQQVDKICEWKCTYSRPNKIRWYKDKBIFS 843
QY 1393 SS---KYRMEAVGCTTRRLVVQQAQADAGEYSCAGGQRLSFSLDVAEPKVVFAKEQPVH 1449
DB 844 GGLKYLIVIEKNVCT--LIINNPEVDTGKYTCANGVPVTHAQTLVLEPPMKYSFLNPLP 901
QY 1450 REVQQAQASATLSCEVAQAQTEVMYKDKGKLSFSK---VRMEAVG-CTTRRLVVQQAQ 1505
DB 902 NTQEIYRTKQAVLTCKVNTPRAPLVWYRGSKAIOEGDPRFIEKDAVGRCT--LTIRKEVE 959
QY 1506 QADAGEYSCAGSQRSL-FHLHVAEPKAVFAKEQASREVQABAGTATLSCEVAQAQTE 1564
DB 960 EDDQAEWTARITQDVFSKVQVYVEEPRHTFVPMKSKQVNESDL---ATLETVDNDKDAE 1016
QY 1565 VTWYKDKKLS-SSSKVRMEAVGCTTRRLVVQQAQADAGEYSCAGGQRLSFHLHVAEPK 1623
DB 1017 VVMHDKGRDIDGKPKVSSNNRKLIIINGARIEDHGEYKCTTKDRTWAQI----- 1070
QY 1624 VVFAKEQ--PAHREVQABAGASATLSCEVAQAQTEVTWYKDKGKLS--SSKVRVEAVGC 1679
DB 1071 IVDAKNKFIKALVDTEVIEKDDVTLMQTKDTKTPGIWFRNGKQISSMPGKPKETQSRNG 1130
QY 1680 TRRLVVOQAQADAGEYSCAGGQRLSFRLHVAELE--POISERPCR-----REPLVVE 1732
DB 1131 THTLKIGKIEMNEADVVEIDQAGLRGSCNVTVLEAEKRPILNWKPKKIEAKAGEPCVVVK 1190
QY 1733 HEDTILTLATLATSATVTWLDQVEIRRSKRH--ETASQDTHLTVLHGAQVLDLSAIVS 1790
DB 1191 PFQIKGTRR-GDPKQAI---LKNKGPIDEMRKLVEVIIKDDVAEIVFKNPQLADTKNA 1246
QY 1791 CRVG-----AGQDF-----PVQVEEVAKFCR 1813
DB 1247 LELGNSAGTALAPFELPVKDKPKPKPLETKNVTABGLDLVWGTDPDDEGAPKVAIIE 1306
QY 1814 LLE-----PVCBELGGTTLACELSPACAEVVR-----CGNTQPRVGRFQWVAEGP 1861

Db 1307 MOEGRSGNWKVGETGTDPKVKDKLKEH-GEYKFRVKALNECGLSDPLTGE---SVLAKNP 1363
Qy 1862 VRSLTVLGLRAEDAGEYVCESDHDSQAQLTVSVPRVVKFMSGLSTVVAEGEATFOCV 1921
Db 1364 -----YGVPGKPKNDALDVKDCHCT-----LAWPEPEDGGAPI--- 1398
Qy 1922 VSPSDVAVVFRDGLLOPSEK-----PAISOGASHSLTISDLVLEDAGQI-TVEABGA 1975
Db 1399 -----TGIIIRERREKSEKMDHQVGGTKPDCCCLTDKKVVEDEKEYLYRVKAVNK 1446
Qy 1976 SSSAALRVREARVLFK-KKLEPQ-----TVBERSVTLVEILT-RPWPELWRTRNA 2024
Db 1447 AGPGCDGDKPKIMKAKKASPEFTGGIKDLRKVGETIKYDVPISGEPLECECLVWVG 1506
Qy 2025 TALAPGKNVHIEHAGARHRLVHNVGFADRGFF-----GCETPDDKTOAKLTVMR 2075
Db 1507 KPLKAVGRVMSSEKHKIMKIENAVRADSGKFTITLKNSSGSCD-----STATVTVGR 1561
Qy 2076 QVRLVRGLQAVAREOGTATME-----VOLSHADVDGWSWTRDGLRFQOG 2119
Db 1562 PTPPKGLDIAOVCADG-ATLSWNPDDDDGGDPLTGYIVEAQDMNKGYIEVG----- 1614
Qy 2120 PTHCLAVRGMHTLTLGLRPEDSGLMVFKABGVHTSARLVVTELPVSPSRPLQ-DVWTT 2178
Db 1615 -----KVDPTTTLKVNGLR--NKGNYKFRVKAVNEGE-----SEPLSADQYTO 1657
Qy 2179 EKEKVTLECELSRP-----NVDVRW--LKDG-----VELRAGKTMAIAAOGAC 2219
Db 1658 IKDPWDEPKGPRPEITDPADRIDIAWEPHPKDGAPIEYIEVVRDPTK----- 1709
Qy 2220 RSLTYRCEFAQGVVYCDADHAQASASVKVQGRYTYLLYRVLAEDAGEIOFVAENAES 2279
Db 1710 -----EWKE-----VKRVPDTNASISGLKEGKEYQFRVRAVNAKGPQ-----PSEPS 1752
Qy 2280 RAQL-RVKELPVTLRPLRDKIAMKHGVLSCQVSRASA-----QVRWPKG 2325
Db 1753 EKQLAKPFIAPWL-----KXNKLKSIITVKGATVWEVKIGGEPFVFNWFKG 1801
Qy 2326 SQELOQPKYELVSD-----GLYRKLIISDVHADE-----DT 2358
Db 1802 NQOLENG--IQLTIDTRKNEHTILCIPSAMRSDVGEYR-LTVKNGHGADEEKANLTVLDR 1858
Qy 2359 YTCADGVKTSQAFFVEQSITVRLQDVTWEP-----APAWFEC----- 2400
Db 1859 PSKNGPLEVSDVF---EDNLNLNWKPPDDDGGEPIEYVEVEKLDATGRWVPCAKVKDT 1915
Qy 2401 -----KAHIDGLKGGQTYQFRVKAVNKEGASDALSTDKTKAKNPYDEPKGTGTPDVVDWADRV 1975
Qy 2408 -----RPPKWLKGTVLQ-----AGGNVGL-----QEGTVHRLMLRR 2440
Db 1976 SLEWEPKSGDGPITQYVIEKKKGHRDQWCEGKVGSGDQTNAEILGLKEGEYQFRVKA 2035
Qy 2441 TCSTWGPVHFTVGSRSARLVVSDIPVVLTRPLEPKTGRLOSVL-----SCDF--- 2492
Db 2036 V--NKAGP-----GEASDPKSRVAVK-----PRNLKPMIDREAMKTIITIKVGNDFEVDV 2083
Qy 2493 ---RPAKAVQVKDDTPLSPSEKFKMSLEGQMAELRLRLMPADAGVYRCA-----GSA 2545
Db 2084 VRGEPKKK-EWIFNEKPDV-DQKRIESEDYKTRFVLRGATRKHAGLYTLTATNASGD 2141
Qy 2546 HSSTEVTVEARVTVTGPLQDAEATEEGWASFSCELSHDEDEVEWSL-----NGMPLYNDS 2601
Db 2142 KHSVEVILGKSPSPGLPLEVSNVYE-----DRADLEWKVPEDDGAPI---D 2186
Qy 2602 PHEISHK-----GRRHTLVILKSIORADAGIVRASLLKYSTARSLEVRVKPVFLK 2651
Db 2187 HYEIEKMDIATGRWVPCGRSET-----TKTVPNLQPGHEYKFRVRAVNBEG 2233
Qy 2652 ALDDLASAE-----ERGTALQCEVSDEPAHVVRKDGVLQ----- 2687
Db 2234 ESDPLTTNTAILAKNPYEVFGKVDKPEL--VDMDKOHVDLAWNAPDDGGAPIEAFVIEKND 2292

Qy 2688 -----PSD-----KYDFLHTAGTGLVVDVSPEDAGLYTCHVGSE 2723
Db 2293 KNGRWEELVVPDOKTATVFNLKEGEYQFISARNKAGTGDPSDPS-----RVVAK 2346
Qy 2724 ETRARVRVHDHLVGTITKRLKTMVELEGESCSPECVLSHESASDPAMWTVGGKTVGSS--- 2780
Db 2347 PRNLAPRIH-----REDLSOTTVKVGATLKFIVHIDGEPADP-VTWSFNGKIGESKQAQ 2399
Qy 2781 -----SRF-----QATRQGRKIYILVREAAAPSAGEVVFVSRGLTSK-----ASLIVR 2823
Db 2400 IENEPYISRFALPKALRKQSGKYTITATNINGDTSVTINIKVKSKPTPKGPIEIVTDVFE 2459
Qy 2824 ERPAALIKPLEDQ-----WV----- 2838
Db 2460 DRATLDWKPDPEDDGEPIEFYIEKMNKTDGIVPCGRSGDTHFTVDSLNGKGDHYKFRVK 2519
Qy 2839 -----APGEDVELRCEL-----SRAGTP-----VHM-----LKDRKA----- 2865
Db 2520 AVNSEGSPDPLETETDILAKNPFDRPRGRPEPTDWDSDHVDLKWDPPLSDGGAPIEBY 2579
Qy 2866 -IRKSQYD-----VVCBGTWAMLVIRGASLKDAGEYTCV-----EASKSTA 2907
Db 2580 QIEKTKYGRWEPAITVPGGOTTATV--PDLTPNEEYBFRVAVNKGSPSDPSDASKAVI 2637
Qy 2908 SLHVEEKANCFTBELTNLOVEEKGTAFTCKTE-HPAATVTWTR-KGLLELRASGKHQPSQ 2965
Db 2638 AKPRNLKPHIDRDALKNLTIKAGQISFDVPVSGSPAPTVTWHPDNREIRNGRVKLDN 2697
Qy 2966 EGLTLRLTISALEKADSDTYTCDIQAOQSRQALLVQGRVHIIE-----DLEDVD- 3015
Db 2698 PEYOSKLVVQMERDGSOTFTIKAVNAGEDBATV---KINVIDKPTSPNGPLDVSVDHG 2754
Qy 3016 -----VQEGSAT-----FRCR 3027
Db 2755 DHVTLNWRAPDDGGIPIENVVIEKYDTASGRWVPAKAVAGDKTTAVVDGLIPGHEYKFR 2814
Qy 3028 ISPANYE--PVHWF--LOKTPL-----HANELNEID-----AOPG----- 3059
Db 2815 VAAVNAEGESDPLETFTGTTLAKDPFKGKTNAPEITDMDKHVDLEWKPPANDGGAPTE 2874
Qy 3060 -----YHVLTLRLA---LKDSGTIYF-----EAG-----DQRASAA 3088
Db 2875 EYVEMKDEFPFNDVAHVAPAGQTNATVGNLKEGSKYEFRIAKNKGALGDPSPSASAV 2934
Qy 3089 LRVTKEPSVFSR-ELTDATITEGEDTLVTCETSTCDIP-MCWTDKGKTLRGSARCQLSHE 3146
Db 2935 AKARNVPPVIDRNSIQEIKVKAGQDFSLNIPVSGEPTPTITWTPEGTPVESDDRMKLNE 2994
Qy 3147 GHRAQLLITGATLQDSGRY--KCEAGGACSSSIVRVA-----RPVRFQBALKOLEV 3196
Db 2995 DOKTRFHVKRALRSDTGYIICAENENGTDTAEVKVTVLDHPSPSPRGPLDVTNIVKDGCD 3054
Qy 3197 L-----EGGAATLRCVLSV-AAPVKW-----CYGNN-----VLR 3225
Db 3055 LAWKEPEDDGGABEISHYVIEKODAAATGRWTAGCESKOTNFHVDLDTQGHYEYKFRVAVNR 3114
Qy 3226 PGDKTSLRQEGAML-----ELV-----VRNLRPQDSGRYSCSFGDQTT 3263
Db 3115 HCDSDPLEAREAIIAKDPFDRADKPGTPEIVDMDKHADLKWTPPADDDGAPIE-GYLVE 3173
Qy 3264 SATLTVTALPAQFICKLRNKEGATATLRCELSKTAPEVWRKGETLRDGRYCLROD 3323
Db 3174 MRPSPGDWVPAVTG-----AGE-LTATV-----DGLKPGQTY----- 3205
Qy 3324 GAMCELOIRGLAWDWAABEYSCVCGEERTSASLTIRMPAHPIFGLRLH-----QESTEGAT 3378
Db 3206 -----QFRVKNALNAGE-----STPDDPSRTWVAKPRHLAPKRNDRMFVQRVKAGQT 3253
Qy 3379 AT--LRCELSKAAPVEW-RKGRESLRDGRHSRLRQUGAVCELIQICGLAVADAGEYSCVC- 3434
Db 3254 LNFVNVVEGEPAPKLEWFLNGSPSSGNGNTHDNNNTKLTITKSTARADSGKYKIVAT 3313

Db 5128 LEDGSKITGYNVEIREYSGTLWTVASDY - NVREBEFTVDKLRBENDYEFVRVAINAAGK 5186
Qy 5257 -----GOLPQVVEELRLQVAPGTRLAKFQKVKGYGAPRLYFPGQGP 5300
Db 5187 GIPSLSPGPIKIQBSGGSRPQIVKPED - TAQPNRRRAVFTCEAVGRPEPTARMRLNGRE 5245
Qy 5301 LTASAHIRMTGKILHLEIISVTRDSGOVAAVYISNAGAAYSARILLVRGDPPEBKP 5360
Db 5246 LPESRYRFEASDGVYKTKIEMWDIDAGEYTVESNPGSDTANLVVQAPPVIEK-- 5303
Qy 5361 ASDVHEQLVPPRMRLRFTPKVKKGSSITFSKVEGRPVPTVHML---REBAERGVLWIG 5417
Db 5304 ---DVPNTILP-----SGDLVRLKITYPSG-TAPFRHSLVLRNEE-----ID 5340
Qy 5418 PDTGTYVASSAQOHSVLVLDVGRHQGTYYTCIASNAGQALCSASLHVSGLPKVEQEK 5477
Db 5341 MDHPTIRIVEDDHILITIPALSREAGRYEYTVNSDGEATTGFWLNVVTGLPEAPQ--- 5397
Qy 5478 VKEALISTFLOTTQAIQAQGLETPASPADLGQRKEEPLAAKEALGHLSLAEVGTETFLQ 5537
Db 5398 -----GPHIISNIGPSTATL 5412
Qy 5538 KLTSQITEMVSAKITQAKLQVPGSDSDSKTPSAPRHRGSRPSSIOESSSESDGDA 5597
Db 5413 SWRPVPTD-----GGSKITSYVVEKRDLSKD--- 5438
Qy 5598 RGEIPDIYVVTADYLPGLAEQDAITLRGQVVEVLDAAPRLWLVRKPTKSSPSRQGWV 5657
Db 5439 -----EWVTTSNVKDMNYIYT----- 5455
Qy 5658 SPAYLDRLKLSPEWGAEAPEFFGAEVSEDEYKARLSSVIOELSSSEQAFFVEELQFLQS 5717
Db 5456 -----GLPE-----NHEYEFVSA----- 5469
Qy 5718 HHLQHLRCPHPVIAVAGQKAVIFNRVDRIGRHFSSFLQELQOCCDTDDVAMCFIKNOAA 5777
Db 5470 -----QENGIG----- 5476
Qy 5778 PEQYLEFLVGRVQAEVSVVSTAIQBYKYKAEALLAGDPGPPPPHLYLEQBPVERVQ 5837
Db 5477 -----APLVS-----EHPHIIARLPDPPTSPPL----- 5498
Qy 5838 RYQALLKELIRKARNRNCALLEQAYAVVSALPQAEKHLVSLMENYPGTLGALGPPI 5897
Db 5499 -----NLEIVQGGDYVTLWSQR-----PL 5518
Qy 5898 RQGHFIVWEGAPGAPMPKGNHRHVPFRNLHVICPRRDRSDRTVSVYFRNMKLSISD 5957
Db 5519 SDG-----GGRL-----RGYIV---EKQEBEHDFWFRCNQNPSPNNYN 5554
Qy 5958 LNDQVEGDDRAFEVWQEREDSVRKYLQARTAIKSSWVKICGIGQRL--ALPVWRPDP 6015
Db 5555 VPNLIDG-----RKY--RYRVFVAVNDAGLSDLAELDQTLFQASGSGEGPK 5597
Qy 6016 FEEELADCTAELGETVLCACRTGTPKPVISYKDGKAVQVDPHILLIEDPGSCALLILD 6075
Db 5598 IVSPLSDLNBEVGRVCTFECSISGSPREYRWFPGKCELVDTSKYTLINKDGQ--VLIIIN 5656
Qy 6076 SLTGVDGQYCMFAASAAGNSTGLKILVQ-----VPPRFVNVKVRAPFFVEGEDAQFTC 6129
Db 5657 DLTSDDADEYCRATNSSGSTRANLAIKTKPRVFIIPKHYGGYEAQ--KGETIELKI 5713
Qy 6130 TIEGAPYQIRWYKDGALLTTGNKFTQSLSEPRSGLLVLIVRAASKEDJGLYECELAVNRLG 6189
Db 5714 PYKAYPQGEARWTKDGEKIENSKFSITTDK--FATLRISNASREDYGEYRVVSVNSVG 5771
Qy 6190 SARASAEILRIOSPMLQAQEQCHREQLVAVEDTTLERADQEVTVLKLGLGKAPGPSTG 6249
Db 5772 SDSGTVNVTV----- 5781
Qy 6250 DLTGPGCPRGAPALQETGSPPTVTGSEAPAVPRVP--QPLLHEGP--BOEPEAI--- 6302
Db 5782 -----ADVPE-PPRFPIIENILDEAVILSWKPPALDGG 5813

Qy 6303 ARAQEWTVPIR-MEGAAMPGAGTCELLWDVHSHVUVRETTQRTYTYTQAID-THTARPPSMQ 6360
Db 5814 SLVNTYTIKREAMGGSW-----SPCAKSRYYTYTTIEGLRAGKQYEF 5856
Qy 6361 VTIEDVQAQT--GCTAQFEAIIIEGDPOPSVTVYKDSVOLVDSTRLSQQOEGTYSVLVR 6417
Db 5857 IIAENKHGQSKPCPEPTA--PVLIPGDERKRRRGYD-----VDEQGIKIVRGKGTVSS---- 5905
Qy 6418 HVASKDAGVYTCLAQNTGGQVLCKAEILLVGLGDNEPSEKSHRRKLSHSPYEVKEEIGRG 6477
Db 5906 ---NYDNVYVDIMWQYV-----PQPVEIKHDH---VLDHYDIHEELGTG 5943
Qy 6478 VFGFVKRVQHKGNKILCAAKEIPLRSRTRAQAYR-ERDILAAALSHPLVTGLLDQPETRKT 6536
Db 5944 AFGVYHVRTERATGNNFAAKEFVMTPHESDKETVKEI:QTMVSLRHPHTLVNLVLDLADPEDDNE 6003
Qy 6537 LILILELCSSEILLDRLY-RKGVVTEAEVKYVIOQLVEGLHLSHGVHLHLDIKPSNLM 6595
Db 6004 MVMIIYFPMWGGELFEKVADEHNKMSDEAVEYMQVCKGLCHMHNNYVHLDLKPENIMP 6063
Qy 6596 VHPAREDIKICDFQFQANITPAELQFQYQSPFVPEIIOQNVPVSEASDIWANGVTSYL 6655
Db 6064 TTKRSNELKLDLDFGLTAHLOPKQSVKVTGTGAEEPAAEVAEGKPVGYTDMWSVGLSYI 6123
Qy 6656 SLTSSSPAGESDRATLLNVLEGRVSHSSPMAAH--LSEDAKDFIKATLQAPQAPPSAA 6713
Db 6124 LLSGLSPFGGENDDETLRNVKS--CDNMDDSAFSGISEDGKDFIRKULLADPNTRMTIH 6181
Qy 6714 QCLSHPMFLK-SMPAEEAHFINTKOLK 6739
Db 6182 QALEHPMLTPGNAPGRDSQIPSSRYTK 6208

RESULT 10

TI3931

projectin - fruit fly (Drosophila melanogaster) (fragment)

C:Species: Drosophila melanogaster

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C:Accession: TI3931

R:Daley, J.; Southgate, R.; Ayme-Southgate, A.

J. Mol. Biol. 279, 201-210, 1998

A:Title: Structure of the Drosophila projectin protein: isoforms and implication for pro

A:Reference number: Z17815; MUID:98300339; PMID:9636710

A:Accession: TI3931

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-6658 <DAL>

A:Cross-references: UNIPROT:O76281; EMBL:AF047475; NID:G3337430; PID:G3337431; PIDN:AAC2

C:Genetics:

A:Gene: projectin

A:Cross-references: FlyBase:FBgn0005666

A:Map position: 4

A>Note: intron positions not resolved (incomplete sequence)

C:Keywords: muscle

Query Match 3.4%; Score 1392; DB 2; Length 6658;

Best Local Similarity 18.4%; Pred. No. 1.3e-32;

Matches 1567; Conservative 1067; Mismatches 2994; Indels 2902; Gaps 382;

Qy 182 YEVRAPENPLGNASAAALVVDSDAADTASRPGTSTALLAHLQRREARREGAPASPPS 241

Db 2 FRVAVNRAG-----PGEPSDKTSIIAKRCRVKPKFVIGE----- 36

Qy 242 TGTCTCTVEGKHARLSCYVTGEPKPEVTWKKDQGLVTEGRRHVVYEDAQNFVLKILFC 301

Db 37 -GLKNVTVKQOTIRFDIKYDGEPEPAATWKGTDNLKFDNQRICLDQLERNSSITIKS 95

Qy 302 KQSDRLGYTCTASNLVGGQTVSSVLVYVRE-----PAVPFKKQLQDLEVREKSAFTLCRVP 357

Db 96 VRKDTGKYKLVLSNSSGTSIESEAQVWVLDRLPLPGGPPEPE-----BIR 139

Qy 358 QPSTAAWFKETELWASAKYIGEEGTERRLTVRNVSADDDAVICETPEGSRTVAELA 417

QY 32 QIVGNPTQVSWEKDQQPVTAGARFRLAQ-----DGDLYRLTILDALGDSGQY 80

Db 2241 PGPIPPVRIESSSTVAEGQTLDLSCVVGQAHAQVTVYKRG-----GSLPARHQVRG--S 2294
Qy 1038 RLVVQOAGKTADGYSCEA-RGQVSRFLHIT-----EPKRMFAKE 1077
Db 2295 RLYIFQASPADAGQVVCRASNGMEASITVVTGTGQANLAYPAGSTQPIRIE----- 2347
Qy 1078 QSVNEVQAEAGASAMLSCEV-AQAOTEVTVYKDGKLLSSSKVGMVKGCTRRLLVLPQA 1136
Db 2348 -----SSSQVAEGQTLDLNCVVPQGSHAQVTVHKRG-----GSLPVHQTGSLRLLY--QA 2397
Qy 1137 GKADAGEYSCBAGGORVSFHLHI---TEPK-----GVFAKEQSVHNEVQAEAGTTAMLS 1187
Db 2398 SPADSGEYVCRVLGSSVPLEASVLVTIEPAGSPALGVTPVRIESSSSQVAEGQTLDLN 2457
Qy 1188 CEVA-QPQTEVTVYKDGKLLSSSKVMEVKGCTRRLLVVOQVGKADAGEYSC-----EAGG 1242
Db 2458 CLVGAQHAQVTVHKRGSLPA-----RHQVHGSRLRL--QVTPADSGEYVCRVVGSSGT 2511
Qy 1243 QRVSFQLHITPKAVFAKEQLVHNEVRTEA-----CASATLSCEVA-QAOTEVTVYKD 1294
Db 2512 QEASVLVTI-QORLSGSHSQGVAVPVRIESSASLANGHTLDLNCVASQAPHTITWYKR 2570
Qy 1295 GKLLSSSKVRIEAGCMRQLVVOQAGADAGEYTC-----AGQRLSFHLDV-----SE 1345
Db 2571 GGSLLPS-----RHQIVG--SLRLIPQVTPADSGEYVCHVNGAGSRETSLVTIQSGSSH 2624
Qy 1346 PKAVFAKEQLAHRKVQABAGAIATLSCEVA-QAOTEVTVYKDGKLLSSSKVMEAVGCT 1404
Db 2625 VPSVSPPIRIESSSPTVVEGQTLDLNCVVARQPAIITWYKRGGSLLPS-----RHQTHGSH 2680
Qy 1405 RRLVVQACQADTGEYSCEAGQRLSFSLDVAEPKVFEA-----KEOPVHREV 1452
Db 2681 LRL--HOMSVADSGEYVCRAN-----NIDALEASIVISPSGSPAPGSSPIRIES 2733
Qy 1453 QAQ---AGASTLSCEV-AQAOTEVTVYKDGKLLSFSSKVRMEAVGCTRRLLVVOQAGQAD 1508
Db 2734 SSSHVAEGETDLNCVVPQAHAQVTVHKRGSLPSHHQTR-----GSLRLHLHVSAD 2787
Qy 1509 AGEYSC-----EAGSORLSFHLHVAEP-KAVFAKEQAPASREVOAEAGT 1550
Db 2788 SGEYVCRVMSGSGPLEASVLVTIEASG---SSAVHVPAPGAGPPIRIEPPSSRV-AE-GQ 2842
Qy 1551 SATLSCEV-AQAOTEVTVYKDGKLLSSSKVMEAVGCTRRLLVVOEQAQADAGEYSC-- 1607
Db 2843 TLDLKCVPQAHAQVTVHKRGSLPARKQVH-----GPLLRL--NQVSPADSGEYSCQVT 2896
Qy 1608 --AGDQRLSFHLV--ABPKVVFAR--EQPAHREVQAE--AGASATLSCEV-AQAOTEV 1657
Db 2897 GSSGTLEASVLVTIEPSSPGPIAPGLAQPIYIEASSSHVTGQTLDLNCVVPQAHAQV 2956
Qy 1658 TWYKDGKLLSSSKVMEAVGCTRRLLVVOQAGQADAGEYSCAGG-----QRLSFLHVA 1712
Db 2957 TWYKRGGSLLPA-----RHQTHGSQLRL--HLVSPADSGEYVCRASGPGPEQASFTVTVP 3010
Qy 1713 ELE-----POISERCREPLVVKHEHDIILLTATLTPSA-ATVTM-----LKQGV 1757
Db 3011 PSEGSYRLRPVISIDP---PSSTVQQQDASFKCLIHDAAPISLEWKTNRNOLEDNV 3067
Qy 1758 EI-----RRSKRHET-----ASQG-----DTHTLTVHGAQVLDSATYSCRVAEG 1797
Db 3068 HISPNGSIITIVGRPSNHGYRCVASNAYGAQSVNLSVHGPPTVS-----VLPEG 3120
Qy 1798 QDFPVQVEEVAAKFCRLLEPVCBGLGTVTTLACELSPACAEVWRWCGNTPRVKGRFOMV 1857
Db 3121 -----PVMVYGVKAVTLEC-----VSAGEPRSSARWTRI 3149
Qy 1858 AEGPVR-----SLTVLGLRAEDAGEYVCESRDDHTSNQLTVSV-----P 1896
Db 3150 SSTPAKLEQRTYGLMDSHAVLQISAKPSDAGTVVCLAQNALGTAQKQVEVIVDTGANAP 3209
Qy 1897 RVWFMGSLSTVAEAGEEATFCQVWSPSDVAV--WFRDGLALLOPSEKFAISQSGASHL 1955
Db 3210 GAOVQAEAEELTVEAGHTATLRCSATGSPAPTIHWSKLRSPLPWQHL-----EGDTL 3263

Qy 1956 TISDLVLEDAQOITVEAGESSAALRVREAPVLFKKLEP--QTVBERSV-----TLEV 2009
Db 3264 IIPRVAQODSQYTCNATSPAGHA-----EATILHVESPPYATTVPEHSAQAGETVOL 3318
Qy 2010 E-LTRPWPPEL--RWTRNATAL---APGKNVBIHAEGARHRLVHLNVGPDAGFFGCEPDP 2063
Db 3319 QCLAHGTFPLTFQWSRVGSSLPGRATARNELLHFERAAPE-----DSGRVRCRVTN 3369
Qy 2064 DKTOAKUTVENRQVRLVRG-----LOAVEAREQGTATMEV--QLSHADVDS----- 2108
Db 3370 KVGSAEAFAPQL---LVQGPGLSPATSIIPAGSTPTVQVTPQLETKTSIGASVEFHCAVPS 3425
Qy 2109 -----WTRDGLRFQOGPTCHLAVRGPMTHTLISGLRPEDSGMLVFKAEG-----VHTSA 2157
Db 3426 DQGTQLRMFKEG---GQLPPGHSVQDG---VLRIONLDQSCQGTYYCOAHGPMWKAQASA 3479
Qy 2158 RLVVTELP---VSFSRRLQDVVTEKEKVTLEC-ELSRPNVDVRLWLDGVELRAGKTMAI 2213
Db 3480 QLVQALPSVLINIRTSVQTVV--GHAVEFECIALGDGPKQVTVWSKVGHLRPG---IV 3534
Qy 2214 AAQACRSLLTYRCEFPADQGVYVCDADHAQSSASVKQGRTYTLLIYRRLAEDAGEIQFV 2273
Db 3535 QSGGWR---IAHVELADAGQYRCTATNA----- 3560
Qy 2274 AENAESRAQLRVKELPVTILVRPLDKIAMEKHGVLGECQVRSASA-----QV 2320
Db 3561 AGTTQSHVLLLVQALP-----QISMPQ-----EVRVPAGSAAVFPCIASGYPTPDI 3606
Qy 2321 RWFKGSQLOLPGPKVELVSDGLYKRLIISDVHAEDEDTYTCD---GDVKTSAQFFVEE 2376
Db 3607 SWSKLDGSLPPDSRL---NNMLMLPSVRPDAGTYVCTATNRQGVKAFAPHLQVPE 3660
Qy 2377 QSI 2379
Db 3661 RVV 3663

RESULT 12

titin - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
R:Accession: S20901; I46520
R:Label: S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380; PMID:1582406
A:Accession: S20901
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-6805 <LAB>
A:Cross-references: UNIPROT:Q28733; EMBL:X64696
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
R:Label: S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U.;
Nature 345, 273-276, 1990
A:Title: A regular pattern of two types of 100-residue motif in the sequence of titin.
A:Reference number: I46520; MUID:90238553; PMID:2129545
A:Accession: I46520
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 4235-5250 <LA2>
A:Cross-references: EMBL:X17329; NID:g1756; PIDN:CAA35207.1; PID:g930251
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro
C:Keywords: muscle

Query Match 2.5%; Score 1046.5; DB 2; Length 6805;
Best Local Similarity 18.5%; Pred. No. 1.3e-22;
Matches 1394; Conservative 968; Mismatches 2748; Indels 2411; Gaps 346;
Qy 17 KAFVVSVKDATALSCQIVGNPTPOVSW-----EKDQPVTAGAFRLAQDGLVRL 67
Db 476 KLVVVRAGCPRLPAIVRGSPAPKVTWRKVIDNVVRKQVDVLDVTMAF-----L 525

Db 2448 NTATVSWKRPDDGGSEITGYHYVERREKKGLRWVATKTPVSDLRCKVTGQEGNTYEF 2507
Qy 1872 -----AEDAGEY-----VCESRDDHTSAQLTVSVPRVVKFMSGLSTVVAEBEGE 1915
Db 2508 VSAENRAGIPGPSDASNVLVMDKOVAYAGPPSNARVDTDTTKKSASLAWGKPHY--DGL 2564
Qy 1916 ATFOCVSPSVA-VWFRD--GAL-----LQPSKF-----AISOGASHSLTI 1957
Db 2565 EITGYVHEHQVGDWTWKDTGTGALRITEFVFDLHTKEKYNFRISAINDAGVGEPAVI 2624
Qy 1958 SDLVLEDAGQITVAEGASS--SAALVRREAPVLFKKLBPQTVVEBSSVTVLELT-- 2012
Db 2625 PDVEI-----VEREAPDELDALR-----RTLVRAGLSIRIFVPIK 2663
Qy 2013 -RPMPELRWTRNATALAPKNVE-----IHAEGARHR-----LVLHNVGPDADRGFG 2058
Db 2664 GRPAPEVWTOKDINLKTTRANIENTESPTLLIIPECNRYDTGKFWMTIENPAGKSGFVN 2723
Qy 2059 CET-----PDDTKOAKLIVEMROVLVRG-----LOAVEAREQGTATMEVQ 2099
Db 2724 VRVLDTPGVNLNRPDTDKDSVTLHW-DLPLIDGSGRITNYIVEKREATRKSYSTVTK 2782
Qy 2100 LSHADVGSWTRDGLRFOGPTCHLAVRGPMTLTLGLRPEDSGLMVFKAGVHTSARL 2159
Db 2783 -----CHKC-----TKVTGLSEGCEYFFRMAENEGIGEP 2814
Qy 2160 VVTELPVSFRPLQDVVTTTEKVTLECELSRPNNVDVRLWK--DGVBLRAGTKMTAAQ 2216
Db 2815 SETKEPVKASE--APSPDLSNI-MDITKSTVSLAWPKPHDGGSKITGYVIEAQRK 2868
Qy 2217 GACRSLTIRCEFPADQGVVCDADAQSSASVQVGRYTYLIRVLAEDAGEIQFVAEN 2276
Db 2869 G-----SDQWTHITTVKGLECVVRNLTEGEYIT-----FQWMAVN 2903
Qy 2277 AESRAQLR-----VKE--LPVTLVRPLRKIAIEKHGVLQCV--SRASAQVRFK 2324
Db 2904 SAGSAPRESRPVIVKEQTMPELDLRLGIYOKLVIAKAGDNIKVEIPVLRPKPTVTKK 2963
Qy 2325 GSQELQPGPKYELVSDGLYRLKLIISDVAEDEDYTCDAQDVKTSAQPFVBEQSIIVRG 2384
Db 2964 GDQVLKQTVQNVVENTATSTILNISECVSRSDSGPYPLTAKNI-----VGEVGDVITI 3016
Qy 2385 LQDVVMEPAWPAWECETSIISV-----RPPKWLKGTVLQAGNVGLBOEGTV-----HRL 2436
Db 3017 VHDIPGPPTGPIKFD-EVSSDFVTFSEWEP-----ENDGGVPSINYVI 3058
Qy 2437 MLRRTCSMTGPHVFTVGSRSSARLVSDIPVVLTRPLEPKTGRELQSVLSCDFRAP 2496
Db 3059 EMRQTDST-----TWVELATTIVIRTYKA-- 3082
Qy 2497 KAVQWYKDDTPLSPSEKPKMSLEGQMAELRILRLMPADAGVYRQAGSAHSSTEVTVEAR 2556
Db 3083 -----TRLTGVEVQFRVKAQ-----NRYCVGPGIITSASI-VANY 3116
Qy 2557 EVTVTGPLQDAEATEEGWASFCSELSHDEBEVW-----SLNGMPLYNDSFHEISHKGRH 2612
Db 3117 PFKVPGPPTGPQVT-----AVTKDSMTISWHEPLSDGSPILG---YHVERK-BRN 3163
Qy 2613 TLVLKSIORA--DAGIVRASSIKVSTSAARLVV-----KPVVFLKALDLSA 2658
Db 3164 GILWQTVSKALVPGNIFKSSGL--TDGIAYFRVIAENNAKSKPSKPSBPVLALDIPD 3221
Qy 2659 BEERGTLAQCEVSPDEAAHVVRKDGVLQSPDKYDFLHTAGTRGLVHDVSPDAGLYTC 2718
Db 3222 PGK--PIPLNTRHTVTILKAK-----PEYTG----- 3246
Qy 2719 HVGSEETRARVRHDLHVIGITKRLKTMVEVLGSCSFECVLSHESASDPAMWTVGGKTVG 2778
Db 3247 --GPKITSYIVKRDLPNGRWLKANFNSILENE-----FTVSGLTED 3286
Qy 2779 SSSRFQATROGRKYLIVVREA-----PSDAGEVVFSVRLGTLTKASLI-VREERPAALIKP 2832
Db 2832 -----PSDAGEVVFSVRLGTLTKASLI-VREERPAALIKP 2832

Db 3287 AAYEFR-----VIAXNAAGAISSPSEBSDAITCDDVDEAPRILVDFPKDVTILK- 3336
Qy 2833 LEDOMVAPGEDVELURCBLSRAGTP-VHMLKDRKAIRKSQKYDVVCEGTMAMLVIRGASLK 2891
Db 3337 -----AGEAFKLEADVSGRPPTMEWTGDKLEGTKLEIKIADPSTYLINKSSRR 3389
Qy 2892 DAGEYTCVEASKSTASHVEEKANCATEELTNLOV-----BEKGTAVFTCKTEHPAAT 2945
Db 3390 DSGAYIL-----TAT-----DPGGFAKHIFNVKVLDRPDPGPPGLAVSEVTS--KCV 3435
Qy 2946 VTWRKGLLELRASCKHQPQSOEGLRLRLTISALEKADSITYTCDIQCAQSRQALLVQGRRV 3005
Db 3436 LSWLPPLDDGCAKIEHYIVQKRETSRL-----AWTNVASEVQ----- 3472
Qy 3006 HIIEDLEDVDVQEGSSATFCRI-----SPANYEPVHWPDLKTPHLANELNEIDAQPG 3058
Db 3473 --VTKLKVTLLKNGEYIFRVMVNVKVCVGEPLSEPV-----LAVNYPGPPD-PPK 3521
Qy 3059 GYHVLTLRQLAKDSGTIYF-----EAG-----DORASAAALRVTE--KPSVFSRELT 3103
Db 3522 NPEVTTI-----TKDSMVVCMGHPSDGGSEIINIVERRDKAGORWKNKKTVTDLRPK 3577
Qy 3104 DATITEGEDLT-----VCETSTCDIPMC 3127
Db 3578 VSGLTEGHEYEFRIWENNAAGISAPRTSPYKACDAVFKPGPGPNRVLQTSRSSISIA 3637
Qy 3128 WTK-----DGKTLRG-SARCQLSHEGH-----RAOLLITGAT--LQDSGRYK-----CE 3168
Db 3638 WNKPIYDGGSEITGYMVEIALPEDEWKIVTPPAGLKATSYTITNLVNEQYKIRIYAMN 3697
Qy 3169 AGGACSSSIV-----RVHARVPRQEALEKOLEVLEGGATLRCLVSSVAAP---VKM 3217
Db 3698 SEGIGEPALVPGTPKABDRMLPPEILDADLRKLWVIR-ACCTLRLFVPIKGRDPPEVKM 3756
Qy 3218 C--YGNVNLPRGDKYSLRQEGAMLELVNRLRPDQSGRYSCSFGD----- 3260
Db 3757 TREHGESL-----DKASIESSTSYLLIVGNVNRFDGSKYILTVENSSGSKSAFVNVRLD 3812
Qy 3261 -----QTTSATLTVTALPAQFIG--KLBN-----KEATEGATATLURCELSKTAP 3302
Db 3813 TPGPPQDLKVKVTKTSVTLTWDPPLDGGSGIKNYIVEKRESTRKAYSTVATNCHKTS- 3871
Qy 3303 VEWKGSSETLRDGRYCLRDGANCCELQIRGLAMVDAAEYSC-VCGERTSASLTIRPMP 3361
Db 3872 --WK--VDQLQEGSSYFR-----VLAENEYIGIGLPAETAESVAKASERPLP 3913
Qy 3362 AHFTGLRHHQBSIEGATATLURCELSKAAPVEWVRKGRESLRDGRHSRQDGAVALQICG 3421
Db 3914 P--GKI-----TLVDVTRNSVLSWEKPE--HDGGSRL--GYIVEMQSKG 3953
Qy 3422 LAVADAGEYSCVCGEERTSATLT--VKALPAKFTGELRNEEBAVEGATAMLCELSK--VA 3477
Db 3954 -----SKWATCATVKVTEATITGLIOGEYSFVSAQNEKGISDPPRLSPVPIAKDLVI 4008
Qy 3478 PVEWRKGPENLRDGRYILRQEGTRCBLQICGLAMADAGEYLCVCGQERTSATLTIRALP 3537
Db 4009 PPAPFK-----LLFNFTFTVLAGEDLIKIDVPFPG-----RPTPTVT-- 4042
Qy 3538 ARFTEDVKNQEARREGATAVLQCELSAAPVEWVRKGSSETLRDGRYSRQDGTCKELQIRG 3597
Db 4043 -----WHKDDVPLKOTTRVNAESTENSSLSLSIKE 4071
Qy 3598 LAMADTGEYSCVCGQERTSAMLTVRALPIKFTGELRNEEBAVEGATAVLURCELSKNAPVEM 3657
Db 4072 ACREDVGHY-----VVKLSNSAG-----EATETLNAII--LDKPGP-- 4106
Qy 3658 WKGHETLRDGRHSRQDGRARCELQIRGLVAEDAGEYLCMCKERTSAMLTVRAMPSEFI 3717
Db 4107 -----TCPVKMD-----EVTADSIITISWEPPKYD 4130
Qy 3718 EG--LRNEEATEGDTATLWCELSKAAPVEMVRKGHETLRDGRHSRQDGSCELSQIRGLA 3775
Db 4131 GGSINNYIVEKRTSTTTTWOIVSATVARTTIKASRLKTG-----CEYQPR-IA 4178

Db 5874 TVKEDKLEA-----PELDLDSE-----LRKGIIVVRAGSGARHIHFKGRPTPDI 5918
Qy 5742 RNVRDIGHFHS-----SFLQ-ELQOCDDDDVAMCFIKNOAAFEQYLFVGRV-QA 5791
Db 5919 TWSREGEFTDKVQVKGWNTQLSDNCDR-NDAGKVIIVKLENSSGTKTAPVTVKVLDT 5977
Qy 5792 ESVVSTAIQFYKYABEALLAGDPSQPPPLQHYLEQPV-----ERVQYQALLKELI 5847
Db 5978 PGPPQNLAVKEVK-----DSAVLVMEP-----PIIDGAKVRNYVDIKREST 6020
Qy 5848 RNKARNRQNCALLEQAVAVSALPOR-----AENKLHVSLMENYPG 5888
Db 6021 R-----KAYANVSKCNKTKFKVENLTGEGAIYFRVWAENEFVGVPVETVD 6067
Qy 5889 TLEALGEPIROGHFIVME-GAPGARMPKWKHNRHVFLFRNLHLVICKPRDRSDRTDTSVYF 5947
Db 6068 AVKAAEPSPPGKVTLTVDVSQTSASLME-----KPEHDDGSRVLGV- 6110
Qy 5948 RNMKLSIDLNDQVEGDDRAFEVWQEREDSV-----RKYLQARTALIKSW 5995
Db 6111 -----VEMQPKTEKWSVVAESKVCNNAVVTGLSSGHEYQFRVKAYNEKGKS 6156
Qy 5996 VKEICGQORLALPWRPDPFEEELADCTAELGETVKLACRVGTGPKPVIISWYKDGKAVQ 6055
Db 6157 DPRVLGVV-IAKDTQPSKLPFKRYSVQAGEDLKEIPVIGRPRPEIFWVKDGEPLR 6215
Qy 6056 VDPHILIEDPGSCALILDSLTGVDSCQYMCFAASAAAGNS-TLGKILVQVPRFVNKV 6114
Db 6216 -QTRVNVVEATSTILHIKSSKDDFGKTYITATNSAGTATENLSVIVLEKPGPGVPV 6274
Qy 6115 R-----ASPFV--EEDAQFT--CTI-----EGAPYQIRWYKDGAL-----LTTG 6151
Db 6275 RFDEISADFVLSWEPAYTGGCQTSNVIIVEKRDTTTTTHIVSATVARTIKVTKLKTG 6334
Qy 6152 NKFO--TLSEPRSGLL-----VLV-----IRAAKEDLGLVECELVNRL 6188
Db 6335 SEYOFRIYAENRYGKSTLDSKPVIVQVFKPEGPGPTVTSVSRDQMLVQWHEPVDNG 6394
Qy 6189 GSARASAEIRIQSPMLQAOEQCHRELQV-----AAVEDDTL-----ERADQEVTSVLK 6236
Db 6395 GSKVLGYHL-----EQEKNSILWVKVKNKTLIDTKFKTKTGLDGLVEYFKVSAB 6444
Qy 6237 RLLGPKAPGSTGLTGCPGCPGAPALQETGSPVVTGTSBPAPVPRVQPLLHEGPE 6296
Db 6445 NIVG-----IASLAKCPNAPVA-----RDPCDPP----- 6468
Qy 6297 QEPBAIARAQSWTPIRMEGAWPCAG--TGELL-----WDVHSHV-VRET----- 6339
Db 6469 GRPEAIVITRN-NVTLKWKPAYDGGSKITGYIVKEDKLPDGRMKASFTTWLETFEFTVS 6527
Qy 6340 ---TQRTYQAIDHTTA-----RPPSMQVTIED-----VQAQTGGTA 6374
Db 6528 GLVEDQRVEFRVIARNAAGNLSEPSSEGAITARDEIDAPNASLDPKYKDIVVHAGETP 6587
Qy 6375 QFEALIEGDPSPVWYKDSVOLVDST-RLSQOQEGTYSYVLVRHVASKDAGVYTCIAQN 6433
Db 6588 VLEADIRKPIPDVWLVMDKGLELETTARMEIKSTIQKTLVVKDCIRTDGGQYVVKLSN 6647
Qy 6434 TGGQVLKAEALLVLGGDNEPD 6454
Db 6648 VGGTKSLPITVKVLDPRGPPE 6668

RESULT 13

S68235
N:Contains: myosin-light-chain kinase, 210K, nonmuscle - chicken
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 09-Jul-2004
R:Waterston, D.M.; Collinge, M.; Lukas, T.J.; Van Eldik, L.J.; Birukov, K.G.; Stepanova,
FEBS Lett. 373, 217-220, 1995

A:Title: Multiple gene products are produced from a novel protein kinase transcription r
A:Reference number: S68235; MUID:96033976; PMID:7589469
A:Accession: S68235
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1906 <WAT>
A:Cross-references: UNIPROT:P11799; EMBL:X52876; NID:G992992; PIDN:CAA37056.1; PID:G99292
R:Shoemaker, M.O.; Lau, W.; Shattuck, R.L.; Kwiatkowski, A.P.; Matrisian, P.E.; Guerra-S
J. Cell Biol. 111, 1107-1125, 1990
A:Title: Use of DNA sequence and mutant analyses and antisense oligodeoxynucleotides to
activity.
A:Reference number: A37099; MUID:90361738; PMID:2202734
A:Accession: A37099
A:Molecule type: mRNA
A:Residues: 649-1906 <SHO>
A:Cross-references: EMBL:X52876
R:Collinge, M.; Matrisian, P.E.; Zimmer, W.E.; Shattuck, R.L.; Lukas, T.J.; Van Eldik, L
Mol. Cell. Biol. 12, 2359-2371, 1992
A:Title: Structure and expression of a calcium-binding protein gene contained within a c
A:Reference number: A44389; MUID:92236611; PMID:1373815
A:Accession: B44389
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1695-1906 <COL>
A:Cross-references: GB:M88284; NID:G212237; PIDN:AAB53767.1; PID:G212238
A:Accession: A44389
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1750-1906 <CO2>
A:Cross-references: GB:M88283; NID:G211371; PIDN:AAA48647.1; PID:G211372
R:Yoshikai, S.I.; Ikebe, M.
Arch. Biochem. Biophys. 299, 242-247, 1992
A:Title: Molecular cloning of the chicken gizzard telokin gene and cDNA.
A:Reference number: S28227; MUID:93073972; PMID:1444462
A:Accession: S28227
A:Molecule type: mRNA
A:Residues: 1750-1906 <YOS>
A:Cross-references: EMBL:M96655; NID:G212744; PIDN:AAA49083.1; PID:G212745
A:Accession: S78216
A:Molecule type: DNA
A:Residues: 1750-1906 <YOW>
A:Cross-references: EMBL:M96987
R:Olson, N.J.; Pearson, R.B.; Needleman, D.S.; Hurwitz, M.Y.; Kemp, B.E.; Means, A.R.
Proc. Natl. Acad. Sci. U.S.A. 87, 2284-2288, 1990
A:Title: Regulatory and structural motifs of chicken gizzard myosin light chain kinase.
A:Reference number: A35093; MUID:90192792; PMID:2315320
A:Accession: A35093
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 935-1438, 'Q', 1440-1906 <OLS>
A:Cross-references: GB:M31048; NID:G212660; PIDN:AAA49069.1; PID:G212661
R:Guerrero Jr., V.; Russo, M.A.; Olson, N.J.; Putkey, J.A.; Means, A.R.
Biochemistry 25, 8372-8381, 1986
A:Title: Domain organization of chicken gizzard myosin light chain kinase deduced from a
A:Reference number: A25810; MUID:87157587; PMID:3030394
A:Accession: A25810
A:Molecule type: mRNA
A:Residues: 1258-1438, 'Q', 1440-1906 <GUE>
C:Genetics:
A:Introns: 1735/3; 1779/1; 1819/1
C:Superfamily: myosin-light-chain kinase, nonmuscle; fibronectin type III repeat homolog
C:Keywords: alternative initiators; ATP; calmodulin binding; phosphoprotein; phosphotran
F:542-599/Domain: immunoglobulin homology <IMM1>
F:935-1906/Product: myosin-light-chain kinase, 108K, smooth muscle (from 5.5kb transcrip
F:1098-1158/Domain: immunoglobulin homology <IMM2>
F:1451-1708/Domain: protein kinase homology <KIN>
F:1459-1467/Region: protein kinase ATP-binding motif
F:1750-1906/Product: telokin (kinase-related protein KRP) (from 2.7 kb transkript) #stat
F:1808-1869/Domain: immunoglobulin homology <IMM3>

Query Match 2.5%; Score 1012; DB 1; Length 1906;
Best Local Similarity 19.7%; Pred. No. 3.4e-22;
Matches 488; Conservative 318; Mismatches 860; Indels 816; Gaps 76;

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2005, 17:29:13 ; Search time 693 Seconds
(without alignments)
5887.802 Million cell updates/sec

Title: US-10-077-130-5
Perfect score: 41273
Sequence: 1 MDQPGSGAPRFLTRPKAFV.....RNREKRALLYKRNHLAQRV 7968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32134	77.9	6620	Q96AA2	Q96aa2 homo sapien
2	8486.5	20.6	2212	Q8NH3	Q8nh3 homo sapien
3	6902	16.7	1319	Q9HCD3	Q9hcd3 homo sapien
4	6619.5	16.0	1596	Q9HCL6	Q9hcl6 homo sapien
5	6128.5	14.8	1579	Q695L2	Q695l2 mus musculu
6	5358.5	13.0	1040	Q8NH2	Q8nh2 homo sapien
7	3865	9.4	767	Q86YC7	Q86yc7 homo sapien
8	3580.5	8.7	26926	Q8WZB3	Q8wzb3 homo sapien
9	3577.5	8.7	26926	Q10466	Q10466 homo sapien
10	3427	8.3	34350	Q8WZ42	Q8wz42 homo sapien
11	3389	8.2	658	Q8NH4	Q8nh4 homo sapien
12	3313	8.0	7962	Q10465	Q10465 homo sapien
13	3270	7.9	1020	Q8NH0	Q8nh0 homo sapien
14	3182.5	7.7	8081	Q7Z120	Q7z120 caenorhabdi
15	3113	7.5	646	Q8NH8	Q8nh8 homo sapien
16	3110	7.5	1944	Q695L3	Q695l3 brachydanio
17	2793	6.8	551	Q8NH7	Q8nh7 homo sapien
18	2653	6.4	542	Q8NH5	Q8nh5 homo sapien
19	2629	6.4	4162	Q88918	Q88918 gallus gall
20	2486	6.0	6632	1 UN89_CAEEL	001761 caenorhabdi
21	2291	5.6	4039	Q7Z246	Q7z246 brachydanio
22	2273	5.5	5516	Q7Z248	Q7z248 brachydanio
23	2233.5	5.4	3094	Q695L1	Q695l1 homo sapien
24	2203	5.3	3262	Q9EQJ5	Q9eqj5 mus musculu
25	2189.5	5.3	5604	Q8WZ53	Q8wz53 homo sapien
26	2168	5.3	3950	Q7YRF5	Q7yrf5 canis famli
27	2144	5.2	17903	Q7RTL4	Q7rtl4 drosophila
28	2087	5.1	16215	Q9NFS3	Q9nfs3 drosophila
29	2087	5.1	18074	Q917U4	Q917u4 drosophila
30	2072.5	5.0	5636	Q96RW7	Q96rw7 homo sapien
31	2067	5.0	2242	Q9P2P9	Q9p2p9 homo sapien

ALIGNMENTS

RESULT 1

ID	Q96AA2	PRELIMINARY;	PRT;	6620 AA.
AC	Q96AA2;			
DT	01-DEC-2001 (Tremblrel. 19, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	Obscurin.			
GN	Name=OBSCN;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Cardiac;			
RX	MEDLINE=21342081; PubMed=11448995; DOI=10.1083/jcb.200102110;			
RA	Young P.W., Ehler E., Gautel M.;			
RT	"Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor			
RT	protein involved in sarcomere assembly.";			
RL	J. Cell Biol. 154:123-136(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Cardiac;			
RA	Gautel M.S.;			
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: Contains 1 PH domain.			
DR	EMBL; AJ002535; CAC44768.1; -.			
DR	HSSP; P56276; 1FHG.			
DR	GeneW; HGNC:15719; OBSCN.			
DR	InterPro; IPR003961; FN_III.			
DR	InterPro; IPR008979; Gal_bind_like.			
DR	InterPro; IPR007110; IG-like.			
DR	InterPro; IPR003598; IG_C2.			
DR	InterPro; IPR000048; IQ_region.			
DR	InterPro; IPR001849; PH_region.			
DR	InterPro; IPR000219; RhoGEF.			
DR	InterPro; IPR001452; SH3.			
DR	InterPro; IPR001412; tRNA-synt_I.			
DR	Pfam; PF00041; fn3_2.			
DR	Pfam; PF00047; ig; 15.			
DR	Pfam; PF00612; IQ; 1.			
DR	Pfam; PF00169; FH; 1.			
DR	Pfam; PF00621; RhoGEF; 1.			
DR	SMART; SM00060; FN3; 2.			
DR	SMART; SM00408; IGC2; 18.			
DR	SMART; SM00015; IQ; 1.			
DR	SMART; SM00233; PH; 1.			
DR	SMART; SM00346; SH3; 1.			
DR	PROSITE; PS00178; AA TRNA_LIGASE_I; UNKNOWN_1.			
DR	PROSITE; PS50010; DH_2; 1.			
DR	PROSITE; PS50853; FN3; 2.			
DR	PROSITE; PS50835; IG_LIKE; 46.			
DR	PROSITE; PS50096; IQ_1.			

32	2067	5.0	2995	2	Q696W0	Q696w0 brachydanio
33	2048	5.0	5198	2	Q76518	Q76518 caenorhabdi
34	2043.5	5.0	5175	2	Q810L3	Q810l3 caenorhabdi
35	2035.5	4.9	3410	2	Q7TN00	Q7tn00 rattus norv
36	1996.5	4.8	803	2	Q80ZF5	Q80zf5 rattus norv
37	1956.5	4.7	17352	2	Q95YM2	Q95ym2 procambarus
38	1896	4.6	4816	2	Q8T103	Q8t103 bombyx mori
39	1891.5	4.6	8625	2	Q86GD6	Q86gd6 procambarus
40	1837	4.5	4824	2	Q95YM1	Q95ym1 procambarus
41	1763.5	4.3	4796	2	Q9W055	Q9w055 drosophila
42	1736.5	4.2	4796	2	Q9NL88	Q9nl88 drosophila
43	1686.5	4.1	8930	2	Q7KQP7	Q7kqp7 drosophila
44	1680.5	4.1	8648	2	Q7KQP6	Q7kqp6 drosophila
45	1677	4.1	8943	2	Q9V4F7	Q9v4f7 drosophila

[illegible]

QY 2101 SHADVGSWTRDGLRFOQGPTCHLAVRGPMTHTLTLGLRPEDSGLMVFKABGVHTSARLV 2160
DB 2101 SHADVGSWTRDGLRFOQGPTCHLAVRGPMTHTLTLGLRPEDSGLMVFKABGVHTSARLV 2160
QY 2161 VTELVPVFSRPLQDVVTTEKEKVTLECELSRPNDVRWLKDGVELRAGKTMWIAAQAQACR 2220
DB 2161 VTELVPVFSRPLQDVVTTEKEKVTLECELSRPNDVRWLKDGVELRAGKTMWIAAQAQACR 2220
QY 2221 SLTIYRCBFADQGVVYCDHAHQSSASVKVQGRTYTLIYRRYLAEDAGEIQFVAENAESR 2280
DB 2221 SLTIYRCBFADQGVVYCDHAHQSSASVKVQGRTYTLIYRRYLAEDAGEIQFVAENAESR 2280
QY 2281 AOLRVKELPVTILVRPLRDKIAMEXHURGVLECOVSPASAOVRWFKGSQELQPGPKYELVSD 2340
DB 2281 AOLRVKELPVTILVRPLRDKIAMEXHURGVLECOVSPASAOVRWFKGSQELQPGPKYELVSD 2340
QY 2341 GLYRKLIISDVHAEDEDTYTCDAGDVKTSAQFFVEEQSIITIVRGLODVTVMPEPAPWPEC 2400
DB 2341 GLYRKLIISDVHAEDEDTYTCDAGDVKTSAQFFVEEQSIITIVRGLODVTVMPEPAPWPEC 2400
QY 2401 ETSIPSVRPPKWLKGTVLQAGNVGLQEBOGTVHRLMLRRTCSNTWGPVHFTVQKSRSSA 2460
DB 2401 ETSIPSVRPPKWLKGTVLQAGNVGLQEBOGTVHRLMLRRTCSNTWGPVHFTVQKSRSSA 2460
QY 2461 RLIVSDIIPVILTRPLEPKTGRELOSIVLSCDFRPAPKAVQVYKDDTPLSPSEKFKWMSLEG 2520
DB 2461 RLIVSDIIPVILTRPLEPKTGRELOSIVLSCDFRPAPKAVQVYKDDTPLSPSEKFKWMSLEG 2520
QY 2521 QMAELRIURLMPADAGVYRCOAGSAHSTEVTVAREVTVTGPLQDAEATREGWASFSCE 2580
DB 2521 QMAELRIURLMPADAGVYRCOAGSAHSTEVTVAREVTVTGPLQDAEATREGWASFSCE 2580
QY 2581 LSHEDEEVWLSNGMPLYNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTARS 2640
DB 2581 LSHEDEEVWLSNGMPLYNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTARS 2640
QY 2641 EVRKVPVFLKALDLSAEBRGTLALQCEVSDPEAHVVRKDGVLQGPSDKYDFLHTAGT 2700
DB 2641 EVRKVPVFLKALDLSAEBRGTLALQCEVSDPEAHVVRKDGVLQGPSDKYDFLHTAGT 2700
QY 2701 RGLVVDVSPEDAGLYTCHVGSEETRARVRVHDHLVGTITKRLKTMVLEGESCSPECVLS 2760
DB 2701 RGLVVDVSPEDAGLYTCHVGSEETRARVRVHDHLVGTITKRLKTMVLEGESCSPECVLS 2760
QY 2761 HESASDPAMWTVGGKTVGSSSRFQATROGRKYILVVRREAPSDAGEVVVFSVRGLTSKASL 2820
DB 2761 HESASDPAMWTVGGKTVGSSSRFQATROGRKYILVVRREAPSDAGEVVVFSVRGLTSKASL 2820
QY 2821 IVRERPAIIKPLEDQWVAPGEDVELRCELBRAGTPVHMLKDKRKAIKRSQKYDVVCEGTM 2880
DB 2821 IVRERPAIIKPLEDQWVAPGEDVELRCELBRAGTPVHMLKDKRKAIKRSQKYDVVCEGTM 2880
QY 2881 AMLVIRGASLKDAGBYTCEVBASKSTASLHVEEKANCFTBELTNLQVEBKGTAVFTCKTE 2940
DB 2881 AMLVIRGASLKDAGBYTCEVBASKSTASLHVEEKANCFTBELTNLQVEBKGTAVFTCKTE 2940
QY 2941 HPAATVTVWRKGLLELRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGAQSRQALLV 3000
DB 2941 HPAATVTVWRKGLLELRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGAQSRQALLV 3000
QY 3001 QGRRVHIITEDLVVDVQSGSSATFCRISIPANYEPVHWFLOKTPHLANELNEIDAQPGGY 3060
DB 3001 QGRRVHIITEDLVVDVQSGSSATFCRISIPANYEPVHWFLOKTPHLANELNEIDAQPGGY 3060
QY 3061 HVLTLRQALKDSGTIYFEAGDORASALRYTEKPSVFSRELTDAITTEGEDTLVLCETS 3120
DB 3061 HVLTLRQALKDSGTIYFEAGDORASALRYTEKPSVFSRELTDAITTEGEDTLVLCETS 3120
QY 3121 TCDIPMCHWKDGKTLRGSGARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180
DB 3121 TCDIPMCHWKDGKTLRGSGARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180

QY 3181 HARPVRFQEALKDLEVLGGGAATLRCVLSSVAAAPVKWCYGNVLRPGDKYSILROEGAMLE 3240
DB 3181 HARPVRFQEALKDLEVLGGGAATLRCVLSSVAAAPVKWCYGNVLRPGDKYSILROEGAMLE 3240
QY 3241 LVVRNLARPQDSGRYSCSPGDDTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKT 3300
DB 3241 LVVRNLARPQDSGRYSCSPGDDTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKT 3300
QY 3301 APVEWRKGETLRRDGDYCLRGDQAMCELOIRGLAMVDAABEYSCVCGBERTSASITIRPM 3360
DB 3301 APVEWRKGETLRRDGDYCLRGDQAMCELOIRGLAMVDAABEYSCVCGBERTSASITIRPM 3360
QY 3361 PAHFIGRLRHQESTEGATATLRCELSKAAAPVEWRKGRESLRDGRHSLRGDQAVCELOIC 3420
DB 3361 PAHFIGRLRHQESTEGATATLRCELSKAAAPVEWRKGRESLRDGRHSLRGDQAVCELOIC 3420
QY 3421 GLAVADAGEYSCVCGBERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCELSKAPVE 3480
DB 3421 GLAVADAGEYSCVCGBERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCELSKAPVE 3480
QY 3481 WRKGPENLRDGDYCLRGDQAMCELOIRGLAMADAGEYLCVCGBERTSATLTVRALPARP 3540
DB 3481 WRKGPENLRDGDYCLRGDQAMCELOIRGLAMADAGEYLCVCGBERTSATLTVRALPARP 3540
QY 3541 IEDVKNOEAREGATAVLQCELSNAAAPVEWRKGETLRRDGRYSLRDQGTCKELOIRGLAM 3600
DB 3541 IEDVKNOEAREGATAVLQCELSNAAAPVEWRKGETLRRDGRYSLRDQGTCKELOIRGLAM 3600
QY 3601 ADTGEYSCVCGBERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKMAPVEWKKG 3660
DB 3601 ADTGEYSCVCGBERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKMAPVEWKKG 3660
QY 3661 HETLRDGRHSLRGDQAMCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKSFIEGL 3720
DB 3661 HETLRDGRHSLRGDQAMCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKSFIEGL 3720
QY 3721 RNEEATEGDTATLWCELSKAAAPVEWRKGETLRRDGRHSLRGDGRSCELOIRGLAVVDAG 3780
DB 3721 RNEEATEGDTATLWCELSKAAAPVEWRKGETLRRDGRHSLRGDGRSCELOIRGLAVVDAG 3780
QY 3781 EYSCVCGBERTSATLTVRALPARPIEDVKNOEAREGATAVLQCELSKAAAPVEWRKGETL 3840
DB 3781 EYSCVCGBERTSATLTVRALPARPIEDVKNOEAREGATAVLQCELSKAAAPVEWRKGETL 3840
QY 3841 RGDYSLRGDGTCELOIRGLHGLSVADTGEYSCVCGBERTSATLTVRAPQVPFREPQLSLQ 3900
DB 3841 RGDYSLRGDGTCELOIRGLHGLSVADTGEYSCVCGBERTSATLTVRAPQVPFREPQLSLQ 3900
QY 3901 AEEGSTATLQCELSBERTATVWVSKGLOLQANGREPRLQCTAELVLQDLOREDTGEYT 3960
DB 3901 AEEGSTATLQCELSBERTATVWVSKGLOLQANGREPRLQCTAELVLQDLOREDTGEYT 3960
QY 3961 CTCGSQATSATLTVTAAPVRFELRLOHQEVDEGGTAHLCCELSRAGASVEWRKGSLOLPP 4020
DB 3961 CTCGSQATSATLTVTAAPVRFELRLOHQEVDEGGTAHLCCELSRAGASVEWRKGSLOLPP 4020
QY 4021 CAKYQWQDGAABELLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRLOSLQOE 4080
DB 4021 CAKYQWQDGAABELLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRLOSLQOE 4080
QY 4081 TGDITARLCCQLSDAESGAVVQWLKEGVELHAGPKYEMRSQCATRELLTHOLEAKDTGEYA 4140
DB 4081 TGDITARLCCQLSDAESGAVVQWLKEGVELHAGPKYEMRSQCATRELLTHOLEAKDTGEYA 4140
QY 4141 CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADEDFEVSCEVSRAGATGVQWCLQGLPLQ 4200
DB 4141 CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADEDFEVSCEVSRAGATGVQWCLQGLPLQ 4200
QY 4201 SNEVTEVAVRDGRITHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPLQDVQ 4260
DB 4201 SNEVTEVAVRDGRITHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPLQDVQ 4260
QY 4261 LSEGDQASFCQRLSRASQOEARWALGGVPLQANEMNDITVEQGTLLHLTLHKVTLUEDAGT 4320

4261	Db	 LSEGDQAFQCRLSRASQEARWALGGVPLQANEMNDITVBQGTJHLITLHKVTFLEDAGT	4320
4321	Qy	VSFHVGTCSSEBAQLKVTAKNTVVRGLENVEALEGEALFECOLSQPEVAHAHTWLLDDPEV	4380
4321	Db	VSFHVGTCSSEBAQLKVTAKNTVVRGLENVEALEGEALFECOLSQPEVAHAHTWLLDDPEV	4380
4381	Qy	RTSENAEYVFPENGILRHLLLLKNLRPQDSRCRVTFLAGDMVTSAFITVRGWRLEILEPLKN	4440
4381	Db	RTSENAEYVFPENGILRHLLLLKNLRPQDSRCRVTFLAGDMVTSAFITVRGWRLEILEPLKN	4440
4441	Qy	AAVRAGAQAARTCTILSEAVPVGEASWYINGAAVOPDDSDTWTVTADGSHQALLILRSAPPHH	4500
4441	Db	AAVRAGAQAARTCTILSEAVPVGEASWYINGAAVOPDDSDTWTVTADGSHQALLILRSAPPHH	4500
4501	Qy	AGEVTFACRDAVASARLTVLGLPPDPEDAENVVASHSHVTILSWAAPMSDGGGGLCGYRVE	4560
4501	Db	AGEVTFACRDAVASARLTVLGLPPDPEDAENVVASHSHVTILSWAAPMSDGGGGLCGYRVE	4560
4561	Qy	VKEGATGQWRILCHELVPGEPCVVDGLAPGETYRFRVAAPVPGVAGEPVHLPTQTVRLAEP	4620
4561	Db	VKEGATGQWRILCHELVPGEPCVVDGLAPGETYRFRVAAPVPGVAGEPVHLPTQTVRLAEP	4620
4621	Qy	KPVPPQPSAPSRQVAAGEDVSLELEVVAEAGEVIMHKGMERTOPQGRFVVSQGRQOML	4680
4621	Db	KPVPPQPSAPSRQVAAGEDVSLELEVVAEAGEVIMHKGMERTOPQGRFVVSQGRQOML	4680
4681	Qy	VIKFTAEQDQGEYHCGLAQGSICPAANAATFQVALSPASVDEAPQPSLPPEAAQBGDLHLWL	4740
4691	Db	VIKFTAEQDQGEYHCGLAQGSICPAANAATFQVALSPASVDEAPQPSLPPEAAQBGDLHLWL	4740
4741	Qy	EALARKRMSREPTLDSISLPEEDGRSQRLPQEAEEVAPDLSEGSTADELARTGDADL	4800
4741	Db	EALARKRMSREPTLDSISLPEEDGRSQRLPQEAEEVAPDLSEGSTADELARTGDADL	4800
4801	Qy	SHTSSDDESRAGTSPSLVTYLKKAGRPGTSPPLASKVGAPAAPSVKPPQOQOQPLAARVPLG	4860
4801	Db	SHTSSDDESRAGTSPSLVTYLKKAGRPGTSPPLASKVGAPAAPSVKPPQOQOQPLAARVPLG	4860
4861	Qy	DLSTKDLDGPDMDKAAVKIOAAFGYKVRKEMKQOEGPMFSHTFGDTEAOGDALLCEV	4920
4861	Db	DLSTKDLDGPDMDKAAVKIOAAFGYKVRKEMKQOEGPMFSHTFGDTEAOGDALLCEV	4920
4921	Qy	VASKADVRARLKDQVBLTDGRHHHIDQLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV	4980
4921	Db	VASKADVRARLKDQVBLTDGRHHHIDQLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV	4980
4981	Qy	THSACVVVSGSESBAESSSGGELDDAFRRAARLRLHRLFRTKSPAESVDESELFLSADEGPA	5040
4981	Db	THSACVVVSGSESBAESSSGGELDDAFRRAARLRLHRLFRTKSPAESVDESELFLSADEGPA	5040
5041	Qy	EPEPADWQTVREDEHFCIRFEALTEARQAVTRFQEMFATLIGIGVEIKLVEQGRPRVEM	5100
5041	Db	EPEPADWQTVREDEHFCIRFEALTEARQAVTRFQEMFATLIGIGVEIKLVEQGRPRVEM	5100
5101	Qy	CISKETAPVVPVPEPLPSLLTSDAAPVFLTELQNEQVODQGPYPSFDCVWTGQPMPSVRWF	5160
5101	Db	CISKETAPVVPVPEPLPSLLTSDAAPVFLTELQNEQVODQGPYPSFDCVWTGQPMPSVRWF	5160
5161	Qy	KQKLLBEDDHMYNEDQOQGHQLIITAVVPADMGVYVYRCLAENSMGVSSTKAELRVDLTS	5220
5161	Db	KQKLLBEDDHMYNEDQOQGHQLIITAVVPADMGVYVYRCLAENSMGVSSTKAELRVDLTS	5220
5221	Qy	TDYDTAADATESSSYFSAQGYLSRREQEGTSTTDGQLPQVVEELRDLQVAPGTFLAKF	5280
5221	Db	TDYDTAADATESSSYFSAQGYLSRREQEGTSTTDGQLPQVVEELRDLQVAPGTFLAKF	5280
5281	Qy	QLKVKGYPARLYWFKDQGPLTASAHIRMTGKKILHTLEIISVTREDSQYAAIISNANG	5340
5281	Db	QLKVKGYPARLYWFKDQGPLTASAHIRMTGKKILHTLEIISVTREDSQYAAIISNANG	5340
5341	Qy	AAVSSARLLVRGPDPEPEKPADVHBOQLVPPRMLERFTPKVKVKGSSITFSVKVEGRPPV	5400


```
10 Q8NH3 PRELIMINARY; PRT; 2212 AA.
AC Q8NH3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Obscurin (Fragment).
GN Name=OBSCN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Young P., Ehler E., Gautel M.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Gautel M.S.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314905; CAC85752.1; -.
DR HSP; Q9UQH9; I5V2.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00408; IGC2; 10.
DR PROSITE; PS50835; IG_Like; 22.
FT NON TER 1
FT NON TER 2212
SQ SEQUENCE 2212 AA; 241979 MW; B6DFA263EB749465 CRC64;

Query Match 20.6%; Score 8486.5; DB 2; Length 2212;
Best Local Similarity 76.5%; Pred. No. 3.2e-265;
Matches 1692; Conservative 1; Mismatches 2; Indels 517; Gaps 3;

Qy 2734 LHVGITKRLKTMVELEGSCFECVLSHESADPAMWTGKTVGSSSRFOATROGRKI 2793
Db 1 LHVGITKRLKTMVELEGSCFECVLSHESADPAMWTGKTVGSSSRFOATROGRKI 60

Qy 2794 LVVREAPSAGEVVFVSGTSMWLVIRGASLKDAGEYTCVEASKSLHVEE 2913
Db 61 LVVREAPSAGEVVFVSGTSMWLVIRGASLKDAGEYTCVEASKSLHVEE 120

Qy 2854 GTPVHMLKDRKAIRKSKYDVVCGTMAWLVIRGASLKDAGEYTCVEASKSLHVEE 2913
Db 121 GTPVHMLKDRKAIRKSKYDVVCGTMAWLVIRGASLKDAGEYTCVEASKSLHVEE 180

Qy 2914 KANCFTEELTNLQVEEKGTAVFTCKTEHPAATVTRKGLLELRASGKHQPSQEGTLRLT 2973
Db 181 KANCFTEELTNLQVEEKGTAVFTCKTEHPAATVTRKGLLELRASGKHQPSQEGTLRLT 240

Qy 2974 ISALEKADSDTYTCDIGQAQRAQLLVQ-GRRVHIIEDLVVDVQEGSSATFRCRISPAN 3032
Db 241 ISALEKADSDTYTCDIGQAQRAQLLVQAGRRVHIIEDLVVDVQEGSSATFRCRISPAN 300

Qy 3033 YEPVHWFLDKTPHANELNEIDAQPGGYHVLTLRLQALKDSGTIYFEAGDQASALRV 3092
Db 301 YEPVHWFLDKTPHANELNEIDAQPGGYHVLTLRLQALKDSGTIYFEAGDQASALRV 360

Qy 3093 EKPSVFSRELTDATITEGEDTLVCETSTCDIPMCWTGDKTGLRSGARCSLHSHRAQL 3152
Db 361 EKPSVFSRELTDATITEGEDTLVCETSTCDIPVCWTGDKTGLRSGARCSLHSHRAQL 420

Qy 3153 LITGATLQDSGRYKCEAGGACSSSIVRVHARVPRFQEAALKOLEVLEGGAATLRCVSSVA 3212
Db 421 LITGATLQDSGRYKCEAGGACSSSIVRVHARVPRFQEAALKOLEVLEGGAATLRCVSSVA 480

Qy 3213 APVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPDQSGRYSCSPGQTTSATLTVTAL 3272
Db 481 APVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPDQSGRYSCSPGQTTSATLTVTAL 540

Qy 3273 PAQFIGKLRNKEATEGATATLRCELSTKTPAVVWRKGSFETLDRGDRYCLURQDGMCELQIR 3332
Db 541 PAQFIGKLRNKEATEGATATLRCELSTKTPAVVWRKGSFETLDRGDRYCLURQDGMCELQIR 600
```

```
Qy 3333 GLAVWDAAEYSCVCGBERTSASLTIRPMPAHFICGLRHQESIEGATATLRCELSCAAAPVE 3392
Db 601 GLAVWDAAEYSCVCGBERTSASLTIRPMPAHFICGLRHQESIEGATATLRCELSCAAAPVE 660
Qy 3393 WRKGRESLRDGRHSLRQDQAVCELOICGLAVADAGEYSCVCGBERTSATLTVKALPAKF 3452
Db 661 WRKGRESLRDGRHSLRQDQAVCELOICGLAVADAGEYSCVCGBERTSATLTVKALPAKF 720
Qy 3453 TEGLRNEEAVEGATAMLCWCELKVPVWVRKGPENLRDGRYIIRQEGTRCELOICGLAM 3512
Db 721 TEGLRNEEAVEGATAMLCWCELKVPVWVRKGPENLRDGRYIIRQEGTRCELOICGLAM 780
Qy 3513 ADAGEYLCVCGBERTSATLTVRALPARFIEDVKNQEARREGATAVLQCELSAAPVWVRKG 3572
Db 781 ADAGEYLCVCGBERTSATLTVRALPARFIEDVKNQEARREGATAVLQCELSAAPVWVRKG 840
Qy 3573 SETLRDGRYSLRODGTCKCELOIRGLAMADTGEYSCVCGBERTSAMLTVRALPIKFTBGL 3632
Db 841 SETLRDGRYSLRODGTCKCELOIRGLAMADTGEYSCVCGBERTSAMLTVRALPIKFTBGL 900
Qy 3633 RNEEATEGATAVLRCELKMAVPVWVRKGHETLRDGRHSLRQDQAVCELOIRGLVAEDAG 3692
Db 901 RNEEATEGATAVLRCELKMAVPVWVRKGHETLRDGRHSLRQDQAVCELOIRGLVAEDAG 960
Qy 3693 EYLCWCKERTSAMLTVRAMPSKPIEGLRNEEATEGDTATLWCELKMAVPVWVRKGHETL 3752
Db 961 EYLCWCKERTSAMLTVRAMPSKPIEGLRNEEATEGDTATLWCELKMAVPVWVRKGHETL 1020
Qy 3753 RDGRHSLRQDGRCELQIRGLAVVDAGEYSCVCGBERTSATLTVRALPARFIEDVKNQ 3812
Db 1021 RDGRHSLRQDGRCELQIRGLAVVDAGEYSCVCGBERTSATLTVRALPARFIEDVKNQ 1080
Qy 3813 AREGATAVLRCELKMAVPVWVRKGSSETLRDGRYSLRQDGTGRCELQIHGLSVADTGEYSC 3872
Db 1081 AREGATAVLRCELKMAVPVWVRKGSSETLRDGRYSLRQDGTGRCELQIHGLSVADTGEYSC 1140
Qy 3873 VCGBERTSATLTVR----- 3886
Db 1141 VCGBERTSATLTVRALPARFTQDLTKKEASEGATATLQCELSKVPVWVRKGPETLRDGG 1200
Qy 3887 ----- 3886
Db 1201 RYSLKQDGTGRCELQIHDLVSADAGEYSCMCGBERTSATLTVRALPARTEGLRNEEAMEG 1260
Qy 3887 ----- 3886
Db 1261 ATATLQCELSKAAVPVWVRKGLEALRDGDKYSLRQDQAVCELOIRGLAMADNGVYSCVCGQ 1320
Qy 3887 ----- 3886
Db 1321 ERTSATLTVRALPARFIEDMRNQKATEGATVTLQCKLRKAAPVWVRKGPNTLRDGRYSL 1380
Qy 3887 ----- 3886
Db 1381 KQDGTSCELQIRGLVIADAGEYSCICBERTSATLTVRALPARFIEDVRNHEATEGATAV 1440
Qy 3887 ----- 3886
Db 1441 LQCELSKAAVPVWVRKGSSETLRDGRYSLRQDGTGRCELQIRGLAVEDTGEYLCVCGBERTS 1500
Qy 3887 ----- 3886
Db 1501 ATLTVRALPARFIDNMTNQAREGATATLHCELSKVPVWVRKGPETLRDGRHSLRQD 1560
Qy 3887 ----- 3886
Db 1561 TRCELQIRGLSVADAGEYSCVCGBERTSATLTIREATEGATAMLCWCELKVPVWVRKGP 1620
Qy 3887 -----APQPVPRFPLQ 3897
Db 1621 ETLRDGRYSLRQDGTGRCELQIHGLSVADTGEYSCVCGBERTSATLTVKAPQPVPRFPLQ 1680
```

3998 SLOAEGSTATLQCELSSEPTATVVMKSGGLQQLQANGRRPREPLOGCTAEVLQDLQREDTG 3957
Db 1681 SLOAEGSTATLQCELSSEPTATVVMKSGGLQQLQANGRRPREPLOGCTAEVLQDLQREDTG 1740
Qy 3958 EYTCTCGSQAATATLTTAAPVRFRLQHQVDEGGTAHLCCELSRAGASVEMRKGSILQ 4017
Db 1741 EYTCTCGSQAATATLTTAAPVRFRLQHQVDEGGTAHLCCELSRAGASVEMRKGSILQ 1800
Qy 4018 LFPCKAYQWQDGAALVVRGVEQEDAGDNYTCDTGHTQSMASLSVRVPRPKFKTRLQSL 4077
Db 1801 LFPCKAYQWQDGAALVVRGVEQEDAGDNYTCDTGHTQSMASLSVRVPRPKFKTRLQSL 1860
Qy 4078 EQETGDIARLCCOLDAESGAVVQWLKEGVELHAGPKYEMRSQGATRELLIHQLEAKDTG 4137
Db 1861 EQETGDIARLCCOLDAESGAVVQWLKEGVELHAGPKYEMRSQGATRELLIHQLEAKDTG 1920
Qy 4138 EYACVTGGQKTAASLRVT-EPVTTIVRGLVDAEVTADVEDVFSRVSAGATGVQWCLQG 4196
Db 1921 EYACVTGGQKTAASLRVTGEPEVTTIVRGLVDAEVTADVEDVFSRVSAGATGVQWCLQG 1980
Qy 4197 LPLQSNVTEVAVRGRHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPL 4256
Db 1981 LPLQSNVTEVAVRGRHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPL 2040
Qy 4257 QDVQSEGDASFQCRLSRASQEARWALGGVPLQANEMNDITVEQGTLLHLTLHKVTLE 4316
Db 2041 QDVQSEGDASFQCRLSRASQEARWALGGVPLQANEMNDITVEQGTLLHLTLHKVTLE 2100
Qy 4317 DAGTVSFHVGTCSSEAOQLKVTAKNTVVRGLENVLEALGEGEALFEQOLSQPEVAHTWLLD 4376
Db 2101 DAGTVSFHVGTCSSEAOQLKVTAKNTVVRGLENVLEALGEGEALFEQOLSQPEVAHTWLLD 2160
Qy 4377 DEPVRTSENAEVFFENGRLHLLLLKNLRPODSCRVTFLAGDMVTSFALTVR 4428
Db 2161 DEPVRTSENAEVFFENGRLHLLLLKNLRPODSCRVTFLAGDMVTSFALTVR 2212

RESULT 3
Q9HCD3
ID Q9HCD3 PRELIMINARY; PRT; 1319 AA.
AC Q9HCD3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KIAA1639 protein (Fragment).
GN Name=KIAA1639;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT xviii. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro."
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046859; BAB13465.1; -.
DR HSSP; P56276; 1FHG.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:000468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 2.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00408; IGC2; 1.

DR PROSITE; PS50853; FN3; 1.
DR PROSITE; PS50835; IG LIKE; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1319 AA; 142048 MW; FBA43AE17204EF48 CRC64;
Query Match 16.7%; Score 6902; DB 2; Length 1319;
Best Local Similarity 99.8%; Pred. No. 2e-214;
Matches 1319; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
Qy 6648 AMGVISLSTCTSPFFAGESDRATLLNVLEGRVSWSPMAAHLSEDAKDFIKATLQAPQ 6707
Db 1 AMGVISLSTCTSPFFAGESDRATLLNVLEGRVSWSPMAAHLSEDAKDFIKATLQAPQ 60
Qy 6708 ARPSAAOCLSHPWFLKMPAEAEHPIINTKQLKFLARSRWRQSLMSYKSLVMSRIPELL 6767
Db 61 ARPSAAOCLSHPWFLKMPAEAEHPIINTKQLKFLARSRWRQSLMSYKSLVMSRIPELL 120
Qy 6768 RGPDPSPSLGVARHLCDRTGCGSSSSSSSSDNELAPPARAKSLPPSPVTHSPHLPGRFLR 6827
Db 121 RGPDPSPSLGVARHLCDRTGCGSSSSSSSSDNELAPPARAKSLPPSPVTHSPHLPGRFLR 180
Qy 6828 PSASLPBEAEASERSTAPAPPASPEGAGPPAAQGCVPRHVSIRSLFYHQAGESPERGAL 6887
Db 181 PSASLPBEAEASERSTAPAPPASPEGAGPPAAQGCVPRHVSIRSLFYHQAGESPERGAL 240
Qy 6888 APGSRHRPARRHLLKGGYTAGALPGLREPLMEHRVLEEAAREEQATLLAKAPSPFETAL 6947
Db 241 APGSRHRPARRHLLKGGYTAGALPGLREPLMEHRVLEEAAREEQATLLAKAPSPFETAL 300
Qy 6948 RLPASGTHLAPGHSHSLSHDPSPTPRPSSEACEAQLPSAPSGGAPIRDMGHPOGSKOL 7007
Db 301 RLPASGTHLAPGHSHSLSHDPSPTPRPSSEACEAQLPSAPSGGAPIRDMGHPOGSKOL 360
Qy 7008 PSTGGHPGTAQPERPSPDPSWGQAPAPFCHPKQGSAPQEGCSPHPAVAPCPGSPGPGSCK 7067
Db 361 PSTGGHPGTAQPERPSPDPSWGQAPAPFCHPKQGSAPQEGCSPHPAVAPCPGSPGPGSCK 420
Qy 7068 EAPLVSSPFLGOAPPAPAKASPLDLSKMGDITSLGRPKGPGCSSPGSASQASSSQ 7127
Db 421 EAPLVSSPFLGOAPPAPAKASPLDLSKMGDITSLGRPKGPGCSSPGSASQASSSQ 480
Qy 7128 VSSLRVGSSQVGTPEPGPSLDAEGWTOEAEDLSSTPTLQRPQOATMRKFSLGRGQYAG 7187
Db 481 VSSLRVGSSQVGTPEPGPSLDAEGWTOEAEDLSSTPTLQRPQOATMRKFSLGRGQYAG 540
Qy 7188 VAGYGTAFGDDAGGMLGGQPMWARIAMAVSQSEEEBQEEARAEESQEEQOEARAESPLP 7247
Db 541 VAGYGTAFGDDAGGMLGGQPMWARIAMAVSQSEEEBQEEARAEESQEEQOEARAESPLP 600
Qy 7248 QVSARPVVEVGRAPTRSSPEPTWEDIGVSLVQIRDLSDGAADAADTISLDISEVDPAYL 7307
Db 601 QVSARPVVEVGRAPTRSSPEPTWEDIGVSLVQIRDLSDGAADAADTISLDISEVDPAYL 660
Qy 7308 NLSLDYDIKYLPEEFMIFRKVPKSAQPEPPSPMAEEELAEFFPTWPGELGPHAGLEI 7367
Db 661 NLSLDYDIKYLPEEFMIFRKVPKSAQPEPPSPMAEEELAEFFPTWPGELGPHAGLEI 720
Qy 7368 TEESDVEDALLAEAAVGRKKKWSRSLPHFPGRHLPLDEPAELGLRERVKASVEHISR 7427
Db 721 TEESDVEDALLAEAAVGRKKKWSRSLPHFPGRHLPLDEPAELGLRERVKASVEHISR 780
Qy 7428 ILKGRPEGLEKEGPPRRKKPGLASFRLSGLKSWDRAPTFRLSDETVVLGQSVTLACQVS 7487
Db 781 ILKGRPEGLEKEGPPRRKKPGLASFRLSGLKSWDRAPTFRLSDETVVLGQSVTLACQVS 840
Qy 7488 AQPAAQATWSKOGAPLESSRVLISATLKNFQLLTLLVVAEDLVYTCVSNALGTVTT 7547
Db 841 AQPAAQATWSKOGAPLESSRVLISATLKNFQLLTLLVVAEDLVYTCVSNALGTVTT 900
Qy 7548 TGVLRKAERPPSSPCPDIGEVADGVLLVWKPVESGPTVYIVQCSLEGSSWTLASDIF 7607

```
Db 901 TGVLRKAERSSPCPDIGEVYADGVLLVWKPVSYPVTVIQCSSLGGSWTTIASDIP 960
Qy 7608 DCCYLTSKLRGGTYTTRTACVSKAGMPYSSPSEQVLLGGPSHLASEEESQGRSAQPLP 7667
Db 961 DCCYLTSKLRGGTYTTRTACVSKAGMPYSSPSEQVLLGGPSHLASEEESQGRSAQPLP 1020
Qy 7668 STKTFATQTOIRGFSVRCQWCKASGRALAAKIIIPVHPKDKTAVLREYALKGLRHPH 7727
Db 1021 STKTFATQTOIRGFSVRCQWCKASGRALAAKIIIPVHPKDKTAVLREYALKGLRHPH 1080
Qy 7728 LAQLHAAVYLSPRHLVLELCSGPELLPCLAEASYSSESEVKOYLWQMLSATQYLHNQHI 7787
Db 1081 LAQLHAAVYLSPRHLVLELCSGPELLPCLAEASYSSESEVKOYLWQMLSATQYLHNQHI 1140
Qy 7788 LHLDRSENMIITYNLLKVVDLNAOGLSSEKVLPSDKFKDYLETWAPLELGGQAVPQ 7847
Db 1141 LHLDRSENMIITYNLLKVVDLNAOGLSSEKVLPSDKFKDYLETW--ELLEGGQAVPQ 1198
Qy 7848 TDIWAIGVTFIMLSAEVPSSEGGARDLQRLKRLGLVRLSRCYAGLSGGAFAFLRSTLCA 7907
Db 1199 TDIWAIGVTFIMLSAEVPSSEGGARDLQRLKRLGLVRLSRCYAGLSGGAFAFLRSTLCA 1258
Qy 7908 QPWGRPCASSCLOCPWLTTEGPACSRPAPVTFPTARLRFVVRNREKRALLYKRHNLAQV 7967
Db 1259 QPWGRPCASSCLOCPWLTTEGPACSRPAPVTFPTARLRFVVRNREKRALLYKRHNLAQV 1318
Qy 7968 R 7968
Db 1319 R 1319

RESULT 4
Q9HCL6 PRELIMINARY; PRT; 1596 AA.
ID AC Q9HCL6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE KIAA1556 protein (Fragment).
GN Name=KIAA1556;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T, Kikuno R., Nakayama M., Hiroseawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT xviii. The complete sequences of 100 new cDNA clones from brain which
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046776; BAB13382.1; -.
DR HSSP; P56276; 1FHG
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00408; IGC2; 7.
DR PROSITE; PS50835; IG_LIKE; 15.
FT NON_TER 1
FT NON_TER 1596
SQ SEQUENCE 1596 AA; 175279 MW; 88937B0760C4E2EF CRC64;

Query Match
Best Local Similarity 81.7%; Score 6619.5; DB 2; Length 1596;
Matches 1309; Conservative 70; Mismatches 203; Indels 21; Gaps 7;

Qy 2726 RARVRVHDLHVGITKRLKTMELVEGESCSFECVLSHESASDPAMWTVGKTVGSSSRPQA 2785
Db 1 RARVRVHDLHVGITKRLKTMELVEGESCSFECVLSHESASDPAMWTVGKTVGSSSRPQA 60
Qy 2786 TRQGRKYLIVREAAPSDAGEVWPSVRGLTSKASLIIVRERPAALIKPLEDQWVAPGEDVE 2845
```

```
Db 61 TRQGRKYLIVREAAPSDAGEVWPSVRGLTSKASLIIVRERPAALIKPLEDQWVAPGEDVE 120
Qy 2846 LRCELSRAGTVPVHMLKDKRKAIKRSQKYDVVCEGTMMAMLVIRGASLKDAGEYTCVEASKS 2905
Db 121 LRCELSRAGTVPVHMLKDKRKAIKRSQKYDVVCEGTMMAMLVIRGASLKDAGEYTCVEASKS 180
Qy 2906 TASHLVEEKANCPTFEELTNLQVEBKGTAVFTCKTEHPAAVTVWRKGLLELRASGHQHPQ 2965
Db 181 TASHLVEEKANCPTFEELTNLQVEBKGTAVFTCKTEHPAAVTVWRKGLLELRASGHQHPQ 240
Qy 2966 EGLTLRLTISALEKADSDTYTCDIGQAQSRALLVQGRRVHIIEDLEDVDVQEGSSATFR 3025
Db 241 EGLTLRLTISALEKADSDTYTCDIGQAQSRALLVQGRRVHIIEDLEDVDVQEGSSATFR 300
Qy 3026 CRISPANYEPVHPFLDKTPLHANELNEIDAQPGYGHVLTIRQLAKDSGTTFYFAGDORA 3085
Db 301 CRISPANYEPVHPFLDKTPLHANELNEIDAQPGYGHVLTIRQLAKDSGTTFYFAGDORA 360
Qy 3086 SAALRVTEKPSVFSRELTDATITTEGEDTLVCESTCDIPMCWTGDKGKTLRGSARCOLSH 3145
Db 361 SAALRVTEKPSVFSRELTDATITTEGEDTLVCESTCDIPVCWTGDKGKTLRGSARCOLSH 420
Qy 3146 EGHRALQILLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQBAKDLLEVEGGAATLR 3205
Db 421 EGHRALQILLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQBAKDLLEVEGGAATLR 480
Qy 3206 CVLSSVAAAPVKWCYGNVLRPGDKYSLRQSGAMLELVVRNLRPQDSGRYSCSFGDQTSSA 3265
Db 481 CVLSSVAAAPVKWCYGNVLRPGDKYSLRQSGAMLELVVRNLRPQDSGRYSCSFGDQTSSA 540
Qy 3266 TLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAPVWRKGSSETLRDGRYCLRDGA 3325
Db 541 TLTVTALPAQFIGKLRNKEATEGATATLRCELSKAAAPVWRKGSSETLRDGRYCLRDGA 600
Qy 3326 MCELIQIRGLAMVDAAYSCVCGEERTSASLTIRMPAHFIFGLRHQESIEGATATLRCEL 3385
Db 601 MCELIQIRGLAMVDAAYSCVCGEERTSASLTIRMPAHFIFGLRHQESIEGATATLRCEL 660
Qy 3386 SKAAPVWRKGRSLRDGRHSLRQDGAVCELOICGLAVADAGEYSCVCGEERTSATITV 3445
Db 661 SKAAPVWRKGRSLRDGRHSLRQDGAVCELOICGLAVADAGEYSCVCGEERTSATITV 720
Qy 3446 KALPAKTEGLRNEEAEGATAMLCESKVAPVWRKGPENLRDGRYIILRQEGTRCEL 3505
Db 721 KALPAKTEGLRNEEAEGATAMLCESKVAPVWRKGPENLRDGRYIILRQEGTRCEL 780
Qy 3506 QICGLAMADAGEYLCVCGQERTSATITRALPARFIEDVKNOEAREGATAVLQCELNSAA 3565
Db 781 QICGLAMADAGEYLCVCGQERTSATITRALPARFIEDVKNOEAREGATAVLQCELNSAA 840
Qy 3566 PVWRKGSSETLRDGRYSLRQDGTKCBELQIRGLAMADTGEYSVCVCGEERTSAMLTVRALP 3625
Db 841 PVWRKGSSETLRDGRYSLRQDGTKCBELQIRGLAMADTGEYSVCVCGEERTSAMLTVRALP 900
Qy 3626 IKPTEGLRNEEAEGATAVLRCESKMAPVWRKGHETLRDGRHSLRQDGARCELQIRG 3685
Db 901 IKPTEGLRNEEAEGATAVLRCESKMAPVWRKGHETLRDGRHSLRQDGARCELQIRG 960
Qy 3686 LVAEDAGEYLCMCKERTSAMLTVRAMPSPKFIETGLRNEEAEGTATATLWCELSKAAAPVW 3745
Db 961 LVAEDAGEYLCMCKERTSAMLTVRAMPSPKFIETGLRNEEAEGTATATLWCELSKAAAPVW 1020
Qy 3746 RKGHETLRDGRHSLRQDGRCCELQIRGLAVVDAGEYSCVCGEERTSATITVLRALPARFI 3805
Db 1021 RKGHETLRDGRHSLRQDGRCCELQIRGLAVVDAGEYSCVCGEERTSATITVLRALPARFI 1080
Qy 3806 EDVKNOEAREGATAVLQCELKAAAPVWRKGSSETLRDGRYSLRQDGTRCELQIHGLSVA 3865
Db 1081 EDVKNOEAREGATAVLQCELKAAAPVWRKGSSETLRDGRYSLRQDGTRCELQIHGLSVA 1140
Qy 3866 DTGEYSVCVCGEERTSATITVRAPOVFPREPLQSLQABEGSTATLQCELSSEPTATVWWSKG 3925
Db 1141 DTGEYSVCVCGEERTSATITVRAPOVFPREPLQSLQABEGSTATLQCELSK-VAPVWWSKG 1199
```

QY 3926 GLOLQANGRRPRLOCTAELVLQDLQREDTGEYTCCTGSOATSATLTVTAAVPRFLREL 3985
Db 1200 PETLRDGGYSILKQDGTICELQIHLDSVADAGEYSCMCQERTSATLTIVRALPARFTBGL 1259
QY 3986 QHQVEDEGTAHLCCELSPAGASVWRKSLQFLPCAKYQMVQDGAABELLVRGVEQEDA 4045
Db 1260 RNEEAMEGATATLQCELSKA-APVEMRWKLEALRDGDKYSLQDGAVCCELQIHGLAMADN 1318
QY 4046 GNYTCDTGTQSMASISVVRPDKFKTRLSQLEQETGDIARLCCQLSDAESGAVVOWLKE 4105
Db 1319 GYVSCVCGERTSATLTIVRALPARFIEDMRNOKATEGATVTLQCKLRKA---APVEMRWK 1375
QY 4106 GVELHAGPKYEMRSOGATRELLIHQLEAKDTGEYACVTGGQKTAASLRVTEPVIIVRGL 4165
Db 1376 PNTLKDGRYSILKQDGTSCELQIRGLVIADAGEYSCICEBERTSATLTIVRALPARFIEDV 1435
QY 4166 VDAEYTADEVDVEFSCEVSPAGATGVQWCLQGLPLQSNVEVAVVRDGRHITLRLKGVTP 4225
Db 1436 RNEHATEGATVTLQCELSKAAP--VEMRWKSETLRDGRYSILR-QDGTICELQIRGLAVE 1492
QY 4226 DAGTVSFHLGNHASSAQLTVRAPEVTILEPLQDVOLSEGQDASFOCLRSRASGQEARWAL 4285
Db 1493 DTGEYLCVCGERTSATLTIVRALPARFIDNMVQAREGATATLHCELSKVAPV--WRK 1550
QY 4286 GCVPLQANEMNDITVEQGTLLHLTKVTLLEDAGTVSFHVGTG 4328
Db 1551 GPETLRDGRHSRLQEN-----RLNPGGGGCGSELGSC 1582

RESULT 5
Q695L2
ID Q695L2 PRELIMINARY; PRT; 1579 AA.
AC Q695L2
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Obscurin-MLCK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15185077;
RA Sutter S.B., Raecker M.O., Borisov A.B., Russell M.W.;
RT "Orthologous relationship of obscurin and Unc-89: phylogeny of a novel
RT family of tandem myosin light chain kinases.";
RL Dev. Genes Evol. 214:352-359(2004).
CC -i- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL; AY603754; AAT80900.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR001951; Histone_H4.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig.c2.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser thr_pkinase.
DR InterPro; IPR008271; Ser thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase_AS.
DR InterPro; IPR008266; Tyr_pkinase_AS.
DR Pfam; PF00047; Ig; 2.
DR Pfam; PF00069; Pkinase; 2.
DR ProDom; PD000001; Prot_kinase; 2.
DR SMART; SM00409; IG; 2.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM02220; S_TKc; 2.

DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00853; FN3; 1.
DR PROSITE; PS00047; HISTONE_H4; UNKNOWN_1.
DR PROSITE; PS00835; IG LIKE_2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1579 AA; 171722 MW; 81A60D50A29F6DEA CRC64;

Query Match 14.8%; Score 6128.5; DB 2; Length 1579;
Best Local Similarity 75.9%; Pred. No. 2.1e-189;
Matches 1235; Conservative 89; Mismatches 237; Indels 67; Gaps 18;

QY 6359 MQVTIEDVQQTGTAQFEALIEGDQPQSVTWYKDSVLQVDSTLSQQOEGTYSVLRLH 6418
Db 1 MQVTIEDVQVQGDMAQDFDAVIEGHPPIVTWYKGSTQLTSSARLSQRQDGTYSVLRLD 60

QY 6419 VASKDAGVYTCLAQNTGGQVLCRAELLVLGDNEDPDSEKSHRRKLSHFYEVKEEIGRV 6478
Db 61 VAPHDAGVYTCVANNAGQVLCRAELLVHGGD-KLDAENQVYRKLSHFYDVQEEIGRV 119

QY 6479 FGFVKRVQHKGNKILCAAKTIPLSRTRAQAYRERDILAALSHPLVTGLDQFETKTLI 6538
Db 120 FGFVKRVQHKGNKMFCAAKTIPLSRKTQAYQERDILATLGHPLVTGLDQFETKTLI 179

QY 6539 LILELCSSELLDRLYKGVVTEAEVKVYIQQLVEGLHYLHSHGVHLHDIKPSNIMLVHP 6598
Db 180 LILELCSSELLDRLFKKGVVTEAEVKVYIQQLVEGLHYLHSHGILHLDIKPNILMVHP 239

QY 6599 AREDIKICDFGFAQNTTPAELQFSQVGSPEFVSPETIIQONPVSEASDIWAMGVISYLSLT 6658
Db 240 AREDIKICDFGFAQKITPSEPOYSKYGSPEFVSPETIIQONPVSEGSDIWAMGVISYLSLT 299

QY 6659 CSSPPFAGESDRATLLNVLEGRVSWSPMAAHLSEDAKDFIKATLQAPQAPSAACLSH 6718
Db 300 CSSPPFAGESDRATLLNVLEGRVSWSPMAAHLSEDAKDFIKATLQTPQAPRSTQSLAH 359

QY 6719 PWFILKMPAEAEHFINTKQLFLARSRWORSLSYKSIILVMRSIPPELLRGPDPSPSLGV 6778
Db 360 PWFILKMPAEAEHFINTKQLFLARSRWORSLSYKSIILVMRSIPPELLRGPDPSPSLGV 419

QY 6779 ARHLCRDTGGSS 6838
Db 420 ARHLRGEASGASS 479

QY 6839 SERSTEAPAPSPGEGAPPAAGCVPVPHSVIRSLFTHQAGESPEHGALAPGSRHRHPARR 6898
Db 480 SMPTADA-AVPASQSPAGPPASPSPVPHSVISLSLFYQAGAEAGNKTSGAKRHPARR 538

QY 6899 RHLLKGYTAGALPGLREPLMEHRVLEEARAEQATLLAKAPSFETALRLPASGTHLAP 6958
Db 539 RHLLKGYTAGALPGLREPLMEYSLLEEARAEQASLMTKTSFETALRLPSSSVREVP 598

QY 6959 GHSHLSHSDSPSTPRPSSSEACGEAQLRPSAPSGG----APIRDMGHFQG----SKQLPST 7010
Db 599 GRSHSLD-NPPVTITGSPPEACKQLLP--PSTGLTHETAKDRGHKEGLQSSVPPPM 655

QY 7011 GHHPGTAQPERPSPDSFWQCAPFCHPKQGSAPQEGCSHPHAPVACPFGSPFGSCKEAP 7070
Db 656 SGDSRFQKQEGSSQDSCRGKPASSCHSELGSGQEGCGPPSSQS---LGLSPPQSLKKE 712

QY 7071 LVPSSP-FLGQPCAPPAPAKASPLDSKMGPGDISLPGPKPGCSPSPGASASASSQVS 7129
Db 713 STSCGLFSEQQAAPFTQVSPLLGSEKEPDGSLSEGPVPPVSPSSPGSASQVDA---- 768

QY 7130 SLRVGSSQVGTPEGPSLDAEGMTQEAEDLSDSPTTLQRPQEQATMRKFSLGRGGYAGVA 7189
Db 769 -----SLDTEG-LSEAGDTCDFTPPQRPQEQATTKFSLSESGGYAGVA 812

QY 7190 GYCTFAFGDAGMLCGGPMWARIANAVSQSEEEQEEARAEASQSEEQEARESEPLPQV 7249

Db 813 GYGTFAFGDAGGMLGGPLNARMWAVSQS-----SEEQDEAATSPQPLE 859

Qy 7250 SARVPVEVGRAPTRSSPEPTWEDIGQVSLVQIRDLSGDAEAADTISLDISEVPAYLNL 7309

Db 860 SLGPIAEASGVPLRTSPSLTPEWEVEQVSLVQIRDLSGDAEAADTISLDISEVPAYLNL 919

Qy 7310 SLDYDIKYLPEEFMIFERKVPKSAQPEPSPMAESEE---LAEP-PEPTWMPGELGPHAG 7364

Db 920 SLDYDIKYLPEEFMIFERKVPKSAQPEPSPMAESEE---LAEP-PEPTWMPGELGPHAG 978

Qy 7365 LEITEESES---DVALAALAAVGRKRKMSPSRSLFHPGRHLPLTILVVAEDLGVYTCVSNA 7421

Db 979 LEITEEPEPCDLEALLGANAAGVRKRW-SPSRGLFQFPGRCLSGEEEPVELGLQRVKAS 1037

Qy 7422 VEHLISRLKRGPELKEGEPKPRKPGLASFRSLGLKSWDRAPTEFLRELSDETIVLQGSVT 7481

Db 1038 MAHISRLIKRGPEGPEREGPRKPKAGLASFRSLGLKGRD-----QELSDEAVVLQGSVT 1091

Qy 7482 LACOVSAQPAQAATWSKDGAPLESSRVLISATLKNFOLLTILVVAEDLGVYTCVSNA 7541

Db 1092 LACOVLAQPTAQAATWSKDGVLLESSGHLLISSTLKNFOLLTILVVAEDLGVYTCVSNA 1151

Qy 7542 LGTIVTTGVLKARPSPPCDIGEVYADGVLLVWKPEVSGPYTYIVQCSLEGGSWTT 7601

Db 1152 LGTAVTTGVLKARPSPPRPEVGEYKDAVLLVWKPEVSGPYTYIVQCSLEGGSWTT 1211

Qy 7602 LASDIFDCYLTSLSRGTYTPTACVSKAGMGPYSSPSQVLLGGPSHLASEE-SQG 7660

Db 1212 LASDISDCYLTGKLSRGMYIFRTACVSKAGMGPYSSPSQVLLGGPSHLASEESRG 1271

Qy 7661 RSAQPLSTKTFATQTOIORGRFSVVRQCEKASGRALAANKIIPVHPKDKTAVLREYAL 7720

Db 1272 RPAQLLSTKTFATQTOIORGRFSVVRQCEKASGRALAANKIIPVHPKDKTAVLREYAL 1331

Qy 7721 KGLRHPHQAOLHAAYLSRPHVLVLELCSGPPELLPCLAEASYSSESEVKDYLMWLSATQ 7780

Db 1332 KRLHHPHQAOLHAAYLSRPHVLVLELCSGPPELLPCLAEASYSSESEVKDYLMWLSATQ 1391

Qy 7781 YLHNOHLHLDRSENMIITEYNLLKVVLDLGNAGSLQSEKVLPSDKFYLETWAPELLE 7840

Db 1392 YLHAQHLHLDRSENMMVTEYNLLKVIDLGNAGSLQSEKVPAPENFKDYLETWAPELLE 1451

Qy 7841 GOGAVPQTDIWAIGVTAIFMLSAEYVSSEGDARDLQRLKGLVRLSCYAGLSGAVAF 7900

Db 1452 GOGAVPQTDIWAIGVTAIFMLSAEYVSSEGDARDLQRLKGLVRLSCYAGLSGAVAF 1511

Qy 7901 LRSTLCAQWGRPCASSCLQCPWLTTEGPACSRPAPVTFPTARLRFVVRNREKRRALLYK 7960

Db 1512 LQSSICAQWGRPCASTCLQCGWLTTEGPTGSRPTPTVTFPTVRLRAFVREKRRALLYK 1571

Qy 7961 RHNLAQVR 7968

Db 1572 KHNLAQVR 1579

RESULT 6

Q8NH2 PRELIMINARY; PRT; 1040 AA.

AC Q8NH2; 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE Obscurin (Fragment).

GN Name=OBSCN;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RA Young P., Ehler E., Gautel M.;

RU Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Gautel M.S.;

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ314906; CAC85753.1; -

DR HSP; P56276; 1FHG.

DR InterPro; IPR003962; FNIII subd.

DR InterPro; IPR003961; FN III

DR InterPro; IPR008957; FN_III-like.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003598; Ig C2.

DR InterPro; IPR000048; IQ_region.

DR Pfam; PF00041; fn3; 1.

DR Pfam; PF00047; ig; 1.

DR Pfam; PF00612; IQ; 1.

DR PRINTS; PR00014; FNTYPEIII.

DR SMART; SM00060; FN3; 1.

DR SMART; SM00408; IGC2; 3.

DR SMART; SM00015; IQ; 1.

DR PROSITE; PS50853; FN3; 1.

DR PROSITE; PS50835; IQ-LIKE; 5.

DR PROSITE; PS50096; IQ; 1.

FT NON TER 1

FT NON TER 1040 1040

SQ SEQUENCE 1040 AA; 3D879CA266D7F44 CRC64;

Query Match 13.0%; Score 5358.5; DB 2; Length 1040;

Best Local Similarity 99.4%; Pred. No. 8.1e-165;

Matches 1034; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 4430 WRLEILPLKNAAVRAGAQAARFTCTLSEAVPVGASWYINGAAVQPDSDTWTADGSHQ 4489

Db 1 WRLEILPLKNAAVRAGAQAARFTCTLSEAVPVGASWYINGAAVQPDSDTWTADGSHH 60

Qy 4490 ALLRSQAQPHAGVTTACRDASARLTVLGLPDPPEDEAVHSHSTVTLSWAAPMSD 4549

Db 61 ALLRSQAQPHAGVTTACRDASARLTVLGLPDPPEDEAVHSHSTVTLSWAAPMSD 120

Qy 4550 GGGGLCGYRVEVKEGATGQWRLCHELVPGPCVVDGLAPGETYFRVAAGVPGVAGEPVH 4609

Db 121 GGGGLCGYRVEVKEGATGQWRLCHELVPGPCVVDGLAPGETYFRVAAGVPGVAGEPVH 180

Qy 4610 LPQTVRLAEPKPVPPQSPAPESQVAGEDVSLSEVVAEAGVVIHKGHERIQPGGRF 4669

Db 181 LPQTVRLAEPKPVPPQSPAPESQVAGEDVSLSEVVAEAGVVIHKGHERIQPGGRF 240

Qy 4670 EVWSQGRQOMLVIKGFTAEQDQGEYHCGLAQSGICPAAATFQVALSPASVDEAPQPSLPPE 4729

Db 241 EVWSQGRQOMLVIKGFTAEQDQGEYHCGLAQSGICPAAATFQVALSPASVDEAPQPSLPPE 300

Qy 4730 AAQEGDLHLWEALARKRMSREPTLDSISLPEEDGRSQRLPQAEAEVAPDLSEGYSTA 4789

Db 301 AAQEGDLHLWEALARKRMSREPTLDSISLPEEDGRSQRLPQAEAEVAPDLSEGYSTA 360

Qy 4790 DELARTGDADLSHTSSDDESAGTPSLVTVLKKAGRPCTSPASKVGAAPAPSVKPPQOO 4849

Db 361 DELARTGDADLSHTSSDDESAGTPSLVTVLKKAGRPCTSPASKVGAAPAPSVKPPQOO 420

Qy 4850 EPLAAVRPPLGDLSTKDLGDFSMDKAAVKIOAATFKYKVRKEMKQOEGPMFSHTFGDTEA 4909

Db 421 EPLAAVRPPLGDLSTKDLGDFSMDKAAVKIOAATFKYKVRKEMKQOEGPMFSHTFGDTEA 480

Qy 4910 QVGDALRLCECVVASKADVRAARWLKDGVELTDGRHHHIDQLGDGTCSLLIAGLDRADAGCY 4969

Db 481 QVGDALRLCECVVASKADVRAARWLKDGVELTDGRHHHIDQLGDGTCSLLIAGLDRADAGCY 540

Qy 4970 TCQVSNKFGQVTHSACVVVSGSEASESSSGELDDAFRRARLRHLFRTKSPAEVSD 5029

Db 541 TCQVSNKFGQVTHSACVVVSGSEASESSSGELDDAFRRARLRHLFRTKSPAEVSD 600

Qy 5030 ELFLSADGPAEPPEPADWQTYREDEHFICIRFEALTEARQAVTRFQEMFATLIGIVBIK 5089

Db 601 ELFLSADGPAEPPEPADWQTYREDEHFICIRFEALTEARQAVTRFQEMFATLIGIVBIK 660

```
QY 5090 LIVEQPRRVMCISKETAPVPPBPLPSLLTSDAAPVFLTELQNEQVODGYVPSFDCVW 5149
Db 661 LIVEQPRRVMCISKETAPVPPBPLPSLLTSDAAPVFLTELQNEQVODGYVPSFDCVW 720
QY 5150 TQCPMPVVRWFKDGLLEDHMYMINEQOGHQLIITAVVPADMGVYRCIAENSMGVSS 5209
Db 721 TQCPMPVVRWFKDGLLEDHMYMINEQOGHQLIITAVVPADMGVYRCIAENSMGVSS 780
QY 5210 TKAELRVDLTSDYDTAADAATESSYFSAQGYLS-SRQEGTSTTDEGQLPQVVEELRD 5268
Db 781 TKAELRVDLTSDYDTAADAATESSYFSAQGYLSRQEGTSTTDEGQLPQVVEELRD 840
QY 5269 LQVAPGTRLAKPQLKVGYPAPRLYWKDGOPLTASAHIRMTGKILHTLBIISVTRDS 5328
Db 841 LQVAPGTRLAKPQLKVGYPAPRLYWKDGOPLTASAHIRMTGKILHTLBIISVTRDS 900
QY 5329 GOYAAIYISNAGAAVSSARLLVRGDEPEEKPADSVHEQLVPPRMLERFTPKVKKGSSI 5388
Db 901 GOYAAIYISNAGAAVSSARLLVRGDEPEEKPADSVHEQLVPPRMLERFTPKVKKGSSI 960
QY 5389 TFSVKVEGRPVPTVHMLREAEARGVLWIGPDPPTGYTVASSAQOHSVLVLDVGRQHGTYT 5448
Db 961 TFSVKVEGRPVPTVHMLREAEARGVLWIGPDPPTGYTVASSAQOHSVLVLDVGRQHGTYT 1020
QY 5449 CIASNAAGQALCSASLHVSG 5468
Db 1021 CIASNAAGQALCSASLHVSG 1040

RESULT 7
Q86YC7 ID Q86YC7 PRELIMINARY; PRT; 767 AA.
AC Q86YC7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Similar to titin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043916; AAH43916.1; -.
DR HSSP; P56276; 1FHG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00408; IGC2; 3.
DR PROSITE; PS50835; IG LIKE; 5.
FT NON_TER
SQ SEQUENCE 767 AA; 84732 MW; 30AB316B63964DAA CRC64;

Query Match 9.4%; Score 3865; DB 2; Length 767;
Best Local Similarity 99.5%; Pred. No. 7.4e-117;
Matches 751; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2248 VKVQGRYTYLLYRVLAEDAGEIQFVAENAESRAQLRVKELPVTLVRLPDKKIAMEKHG 2307
Db 5 IQREGRTYTYLLYRVLAEDAGEIQFVAENAESRAQLRVKELPVTLVRLPDKKIAMEKHG 64
QY 2308 VLEQVSRASQVRFWFGSQBLQPKYVELVSDGLYRKLIIISDVHAEDTYTCDAGDVK 2367
Db 65 VLEQVSRASQVRFWFGSQBLQPKYVELVSDGLYRKLIIISDVHAEDTYTCDAGDVK 124
QY 2368 TSAQFFVEQSITIVRGLQDVTVMEPAPWAFECETSIPIVPPKWLIGKTVLQAGGNVGL 2427
Db 125 TSAQFFVEQSITIVRGLQDVTVMEPAPWAFECETSIPIVPPKWLIGKTVLQAGGNVGL 184
QY 2428 EQEGTVHRLMLRRTCTMTGPHFTVGKSRSSARLVVSDIPVLTTRPLEPKTGRELSV 2487
```

```
Db 185 EQEGTVHRLMLRRTCTMTGPHFTVGKSRSSARLVVSDIPVLTTRPLEPKTGRELSV 244
QY 2488 LSCDFRPAPKAVOMYKDDTPLSPSEKFKMSLEGQMAELRILRLMPADAGVYRCOAGSAHS 2547
Db 245 LSCDFRPAPKAVOMYKDDTPLSPSEKFKMSLEGQMAELRILRLMPADAGVYRCOAGSAHS 304
QY 2548 STEVTVEAREVTVGPIQDAEAEVEGWASFCSELSHDEDEVEVSLNGLMPLYNDSEFHISH 2607
Db 305 STEVTVEAREVTVGPIQDAEAEVEGWASFCSELSHDEDEVEVSLNGLMPLYNDSEFHISH 364
QY 2608 KGRHRTLVLKSIQADAGIVRASGLKYVTSARLEVRVKPVVFLKALDDLSAEERGTLLAQ 2667
Db 365 KGRHRTLVLKSIQADAGIVRASGLKYVTSARLEVRVKPVVFLKALDDLSAEERGTLLAQ 424
QY 2668 CEVSDPBAHVVRKOGVOLGSPDKYDFLHTAGTGLVVDVSPEDAGLYTCHVGSEETRA 2727
Db 425 CEVSDPBAHVVRKOGVOLGSPDKYDFLHTAGTGLVVDVSPEDAGLYTCHVGSEETRA 484
QY 2728 RVRVHDLHVGITTKRLKTMEVLEGESCSFECVLSHESASDPAMWTVGGTKVSSSRFOATR 2787
Db 485 RVRVHDLHVGITTKRLKTMEVLEGESCSFECVLSHESASDPAMWTVGGTKVSSSRFOATR 544
QY 2788 QGRKYILVREAAAPSADAGEVVFVRGLTSAKSLIVRERPAALIKPLEDQWVAPGEDVELR 2847
Db 545 QGRKYILVREAAAPSADAGEVVFVRGLTSAKSLIVRERPAALIKPLEDQWVAPGEDVELR 604
QY 2848 CELSRAGTPVHMLKDRKAIKRSOKYDVVCEGTWMLVIRGASLKDAGEYTCVEASKSTA 2907
Db 605 CELSRAGTPVHMLKDRKAIKRSOKYDVVCEGTWMLVIRGASLKDAGEYTCVEASKSTA 664
QY 2908 SLHVEKANCFTBELTNLQVEEKGTAFTCTKSHPAATVTRKGLLELRASGKHQPSQEG 2967
Db 665 SLHVEKANCFTBELTNLQVEEKGTAFTCTKSHPAATVTRKGLLELRASGKHQPSQEG 724
QY 2968 LTURLTISALEKADSDTYTCDIQOASRAQLLVQG 3002
Db 725 LTURLTISALEKADSDTYTCDIQOASRAQLLVQG 759

RESULT 8
Q8WZB3 ID Q8WZB3 PRELIMINARY; PRT; 26926 AA.
AC Q8WZB3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE N2B-Titin Isoform.
GN Name=TTN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309627; PubMed=10850961;
RA Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereousse F.,
RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,
RA Granzier H., Labeit S.;
RT "Series of exon-skipping events in the elastic spring region of titin
as the structural basis for myofibrillar elastic diversity.";
RL Circ. Res. 86:1114-1121(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21573839; PubMed=11717165;
RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,
RA McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,
RA Labeit S.;
RT "The complete gene sequence of titin, expression of an unusual ~700
kDa titin isoform and its interaction with obscurin identify a novel
Z-line to I-band linking system.";
RL Circ. Res. 89:1065-1072(2001).
RN [3]
```


8764	EGQEY---KFRVIAKNKFGCGPPVEIGP---ILAVDP-----	8799
6624	YGSEPFVSPBIIQONPVSEASDIWAMGVISYLSITCSSPFAGESDRATLLNVLEGRVSW	6683
8795	LGPP--TSPE-----RLIYTERTKST-----ITLDWKEPRNGG	8826
6684	SPMAHLSE---DAKDFIKATIQAPQAPRPSAAQCLSHFWFLKSPMAEBAHFINTYKQLK	6739
8827	SPIQGYIIEKRHDKPDF-----BRVKNKLCPTTSFLVENLDEHQMYFRVK---	8873
6740	FLLAARSQRSLMSYKILVMRSIPELLRGPPDPSPIGVARHLCRDTGGSSSSSSSSDNE	6799
8874	-----AVNEIGESEPSLP-LNVV-----TQDDE	8895
6800	LAPFARAK-----SLP---PSPVTHSP-----LHPRGFLRPSASL---PEE	6935
8896	VPPTIKLRLSVRGDTIKVKAGEPVHIADVTGLPMPKIEWSKNETVIEKFTDALQITKEE	8955
6936	AEASERSTEPAPASPPEGAGPPAAQCVPRHVSIRLSLFYHQAGESPEHGALLAPGRRRHP	6995
8956	VSRSEAKTELSPKAVREDKGTVTVTASNRLGSVFRNVHVEVDR-----PSPPRNL	9007
6996	ARRHLLKGGYIAGALP---GLREPLMEHRLVEEERABEQ-----TLLAK-----	6939
9008	AVTDIKAESCULTWDAPLDNGGSE---ITHXVIDKRDASRKKAWEVNTTAVEKRYGIWK	9065
6940	APSFETALRLPASGTHLAPCHSHLSHDSPSTRPSSSEACGEAQRLPSPAGCAPTRDM	6998
9066	LIPNGQYEFVRVANKY---GISDEKSDKVVITDP-----YRLPGPP--GKP-KVL	9111
6999	GHPOGSKOLPST-----GSHPGTAQ--PERPSPDP-WGQAPAPFCHPKQG-----	7044
9112	ARTKGLMLVSWTPPLDNGGSPITGYMLEKREBEGSPYMSVRAPITKVGLKGVEFNVPL	9171
7045	EGCS-PPHVAVAPCPGSPFGSGCKEAPLVPSSP-FLGQQAAPPAPAKASPPLDKXMGPG	7101
9172	LEGVKYQFRAMAINAAGIGPPSE-----PSPDEVAGDPIFFPGP-PSCEPVKDKT-KS	9222
7102	DISLPGRPKPGPCSSPGSASQASSQSVSSLRVSSQVGTPEFPGPSLDAEGW---TOE	7159
9223	SISLGWPP---PAKGG-----SPIKGYIVEMQEET-----TDMKRVNEPDKLI	9264
7160	DS-----TPTLQRPQEQATMRKFLSGRGGYAGVAGYGTAFGGDAGMLQOGPMWARI	7215
9265	TTCECVVPLKE-----LKYRP-----RVK-9285	
7216	AVSQSEEEBQEEARAEQSEEQEAREASPLQVQSAHPVEVG-----RAPT	7262
9286	AVNEAGESPSTDTGEIPATDIOBE-----PEVFI---DIGAQDCLVKAGSQIRIPA	9335
7263	RSSPEPTP---WEDIGOVSLVQ---IRDLSGDAEAADTISLDISEVDPAYNLNSLDYDI	7315
9336	VIGRPTPKSWEPPDGKAKAMKDGVDHIDIEDAQLETAENSSVIIPECKRSHTGKYSIT	9395
7316	KYLPPEPFMIFRKVPKSAQPEPPSPMAEELAEFP-EPTWPWCELG-----	7360
9396	AKNKAGKGTANCRVKVMVDGPPPKOLKVSDITRGSCRLSKMPPDDGGDRIKGVVIEKRT	9455
7361	-----PHAGL-----ETEESEDVDALLAEAAVORKKRWKSSPSRLSFFHPCGRHL	7404
9456	IDGKAMTKVNPDCGSTTFVVVPDLLSEQQYFFRVRAENRFGIGPVPETIQRTTARDP--	9513
7405	PLDEPAELGLRERVKASV-----EHLISRLTKGRPELKEG-----	7440
9514	PPDPPIKLGITKNTVHLVSWPKPKNDGSGPVTHYIVCELANDPTGTCKEAWRQCNKRD	9573
7441	-----PPRKXPLGSLFRLSLGKLSW-----DRAPTF-LRELSDET	7474
9574	VEELOFTVEDLVEGGEYEFVRKAVNAAGVSKPSATVGCPCQRPDMPPSIDLKEFME--	9631
7475	VLGOSVTLACOVSAQPAQAATWSKDGAPLESSSRVLISATLKNQFLL--TILVVVAE---	7529
9632	EEGTNNVIAKIGVPPETLTWFKAPPKPKDNKEPVLVYDTHVKNLVVDVDTCTVILPQSR	9691

```

Qy 7530 -DLGVYTCVSNALGTVTTTGVURKAEKRSSSPCPDIG-----EVTADGVLLVWKPVESY 7580
Db 9692 SDTGLYTITAVNNLGTASKEMRLNVLGRPG----PPVGPIKFESVSADQWTLTSWFPPKOD 9747
Qy 7584 GP---VTYIVQ-CSLEGGSWTTLASIDFDCYLTSKLRSRGTYTFRTACVSKAGMG-PYS 7638
Db 9748 GGSKITNYVIEKEANKRWVHVSSEPKECTYTPKLLLEGHEYVFPRIMAQNKYIGEPD 9807
Qy 7639 SPSE 7642
Db 9808 SEPE 9811

RESULT 9
Q10466 PRELIMINARY; PRT; 26926 AA.
AC Q10466;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Titin, heart isoform N2-B (EC 2.7.1.-) (Connectin).
GN Name=titin;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa;.
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RZ SEQUENCE FROM N.A.
RP RP
RC TISSUE=Heart;
RX MEDLINE=96026330; PubMed=7569978;
RA Labelt S.; Kolmer B.;
RT "Titins: giant proteins in charge of muscle ultrastructure and elasticity.";
RL Science 270:293-296(1995).
RN [2]
RZ SEQUENCE OF 2277-25376 FROM N.A.
RP RP
RX MEDLINE=92258380; PubMed=1582406;
RA Labelt S., Gaucel M., Lakey A., Trinick J.;
RT "Towards a molecular understanding of titin.";
RL EMBO J. 11:1711-1716(1992).
RN [3]
RZ SEQUENCE OF 1976-2014 FROM N.A.
RP RP
RA Labelt S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RZ CHARACTERIZATION.
RX MEDLINE=95331314; PubMed=7607248;
RA Gautel M., Castiglione-Morelli M.A., Pfuhl M., Motta A., Pastore A.;
RT "A calmodulin-binding sequence in the C-terminus of human cardiac titin kinase.";
RL Eur. J. Biochem. 230:752-759(1995).
CC CC
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC CC
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC CC
CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE KINASES.
CC CC
CC -1- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112 IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC CC
DR EMBL; X64698; CAA45939.1; -.
DR EMBL; X90568; CAA62188.1; -.
DR FIR; I38344; I38344.
DR DR
DR DBP; LBPV; NMR; @=14850-14961.
DR PB; IGIC; X-ray; A/B=2027-2125.
DR PB; INCT; NMR; @=26050-26155.
DR PB; INCU; NMR; @=26050-26155.
DR DB; ITIT; NMR; @=5244-5341.
DR PB; LTII; NMR; @=5244-5341.
DR DB; LTKI; X-ray; A/B=24748-25068.
DR GO; GO:0005856; C:cytoskeleton; IEA.
DR GO; GO:0016020; C:membrane; IEA.
```


Qy 1011 AEVMMYKDGKLSLKVHVEAKGRRRLVVQAGKTDAGDYSCEARGQVSRFLHITEP 1070
Db 2428 TSVKWLNDQEI KPPDRVQAIVKGTQKRLVNRTHASDEGPYKLVGVETNCLNSVEKI 2487
Qy 1071 KMFPAKEQSVHNEVQAEAGASAMLSCEVAQAOTEVTWYKDGKLSLSSSKVMGKVGCTRR 1130
Db 2488 KIIRGLRDLTCTEQ-----NVVFEVLSHSGIDLWNFKDKEIKPSSKYIEAHGKIYK 2542
Qy 1131 LVLPOAGKADAGEYSCAGGQVSRFLHITEPQVFAKEQSVHNEVQAEAGTMTAMLSCEV 1190
Db 2543 LTVLNMKDDDEGKTYTFVAGENMTSGKL--TVAGGAISK--PLTDQTVAES--QEAVFCEV 2597
Qy 1191 AOPOTEVTWYKDGKLSLSSSKVMGKCTRRLLVVQVQKADAGEYSCAGGQVSRFLH 1250
Db 2598 ANPSKGEWLDRGKHLPLTNIRSESGHRRLLIAATKLDIDIGEYTYKVTATSKAKL 2657
Qy 1251 ITEPKAVFAKEQLVHNEVRTAGASATLSCEVAQAOTE-VTWYKDGKLSLSSSKVRIBAA 1309
Db 2658 V---EAVKIKKTL-KNLTVTET-QDAVFTVELTHPNVKGQVWIKNGVLESNEKVAISVK 2712
Qy 1310 GCMRLVVQAGQADAGEYTCAG--GORLSFHLD-----VSEPKAVFAKEQLAHRKVOAE 1363
Db 2713 GTIYSLRIKNCALVDESUYGPRGLGASARLHVETVKIIKKPKDVTALEN-----2763
Qy 1364 AGATILSCEVAQAQ--TEVTWYKDGKLSLSSSKVRMEAVGCTRRLLVVQACQADTGEYS 1421
Db 2764 ----ATVAFEVSVSHDTPVPMFKHSVEIKPSDKHRLVSEKRVKHLMLQNTSPSDAGEYT 2819
Qy 1422 CEAGQRLSFLSDVAEPKVPFAKQPVH-----REVQAQAGASTTLCSEVAQAOTEVMMY 1476
Db 2820 AVVQLECKAKL FV-----ETLHITKMKNIEVPETKTASFCEVSHFNVPSMWL 2869
Qy 1477 KDGKLSFSSKVRMEAVGCTRRLLVVQAGQADAGEYSCAGSQRSLFHLHVAEPKAVFAK 1536
Db 2870 KNGVEIEMSEKFIWVQGLHQLIMNTSTEDSAEYTFVCGNDQVSATLTVT-PIMITS- 2927
Qy 1537 EQPASREVOAEAGT SATLSCEVAQAOTEVTWYKDGKLSLSSSKVRMEAVGCTRRLLVQEA 1596
Db 2928 ---MLKDNABEKTITFEVTNVEGYSYKWLKNGVEIKSTDKCOMTKLTHSLNIRV 2984
Qy 1597 QOADAGEYSCAGQRLSFLHVAEPKVPFAKEQPAHREVQAEAGASATLSCEVAQAOTE 1656
Db 2985 HFGDAADYTFVAGKATSTALVVEARHIEFRKH---IKDIKVEKKRAMFECEVSEPDIT 3041
Qy 1657 VTWYKDGKLSLSSSKVRVEAVGCTRRLLVVQAGQADAGEYSCAGGQRLSFLHVAELEP 1716
Db 3042 VQWKKDQDELQITDRIKIQEKYVHRLIIPSTRMSDAGKYTWAGGNVSTAKLFVEGRDV 3101
Qy 1717 QISERPCCREPLVKEHEDIILTATLTPSAATVTLKDGVEI--RRSKRHETASQGDTH 1774
Db 3102 RI--RSIKKEVQVIEKQRAVV--BEFVNEDDVAHWYKDGIEINFPQOERHKYVVERRIH 3157
Qy 1775 TLTVHGAQVLDSALYSCRVAGEGQDFPVQVB-EVAAKFCRLLEPVCGLGTVTLACELS 1833
Db 3158 RMPITSETRQSDAGEYTFVAGNRSSVTLVYNAPPEPPQVLQELQPTVQSG-----K 3208
Qy 1834 PA--CAEVVWRCGNTQPRV-----GKRFQVMAEGVRSILVGLRAEDAGEYVC 1880
Db 3209 PARFCAMI--SGRPQPKISYKBEQLLSTGFKCKFLHDGQEYTLILLIEAFPDAAVTC 3265
Qy 1881 ESRDDH----TSAQITVSVPRVK-----FMSGLTVAEBEGEATFOCVSPS 1925
Db 3266 EAKNDYGATTASLSVEVPVSPQDBMPYVPPAIITPLQDPTVTSEQPAPFCQVSGT 3325
Qy 1926 DVAVVWFRDQALLOPSEKFAISQSGASHSLTISDLVLEDAGQITVEABGA-----SSSAAL 1981
Db 3326 DLKVSWSKDKKIKPSRFRMTQPEDTYQLIEAGAYPEDEGTYTFVANNVAGVQSSTANL 3385
Qy 1982 RVREAPVLFKKLBPQ-----TVRESSVTLLEVELTR-----2013
Db 3386 SLEAPESILHERIEQEIEMEMKEFSSSFLSABEESGLHSAELQLSKINETLELLSESPPV 3445
Qy 2014 -----PWPELWTRNATALAPGKNVEI 2035

Db 3446 TKPSEKEGTGPIPIKEVSNADISMGDVATLSVTVIGIPKPKIQMFFNGVLLTSPADYKP 3505
Qy 2036 HAEGARHRLVLHNVGVFADRGFFGCTPDD-----KT 2066
Db 3506 VFDGDDHSLIILFTKLEDEGEYTCWASNDYKTKICSAYLKINSKGEKHDTETESAVAKS 3565
Qy 2067 QAKL-----TVMEQVRLVRGLQAV-EAREOGTATMEVOLSHADVDGWSWTRDGLRFQ 2117
Db 3566 LEKLGPCPPHFLKELPIRCAOGLPAIFETVYVGEPAPTY-----TWFKENKQL- 3615
Qy 2118 QGP7C---HLAVRGP--MHTLTLSGLRPEPSGLMWFKAEV---HTSARLVV-----2161
Db 3616 ----CTSVYITIIHNPNGSGTFIVNDFQREDSGLYICAEENMLGESTCAABELLVLEBTD 3671
Qy 2162 -----TELPVSFSR-PL-----QDVVTTEKSKVTLECELSRPNVDVRW 2198
Db 3672 MTDTPCKAKSTPEAPEFPQTPLAGPAVEALDSQEIATFVKDTILKAALITEENQQLSY 3731
Qy 2199 -----LKDGVLRAGKTMAIAAQ-----ACRSLTIY-----2225
Db 3732 EHIKANELSQLPLGAEQLOSLILEQDKLTPESTREFLCINGSIHFOPLKEPSNLQLOI 3791
Qy 2226 ---RCEPADQGVVCDADAQOS-----2244
Db 3792 VQSKTFSKEGILMPPEPETQVQLSDTEKIFPFSAMSIEQINSLTVEPLKTLAEBEGNYP 3851
Qy 2245 SASVVOGQTY-----DAGEIQVA---ENAESRAQLRVKE-----2287
Db 3852 QSSIEPPMHSYLTSAVEBLSLKEKTVSDTNRQORVTLQKQEAQSALILSQAELGHVES 3911
Qy 2286 -----LPVTLVRPLDKIAMEKHGV-----2308
Db 3912 LQSPDVMI SQVNYBPLVPSEHSCTEGGKILITESANPLENAGQDSAVRIEESKSLRPLAL 3971
Qy 2308 -----LEKVLKBEHSDNVVMPDQIIESKREPAVKKQVQGRDILSKLSLLSGIPBEQRLN 4031
Db 2322 LECQVSRSA-SAQV-----RW-----2322
Qy 4032 LKIQICRALQAASVASEQPLFSEWLRNIEKVEAVNITQEPRIHMCWLVTSAKSVTEE 4091
Db 2323 -----FKGSGQ 2327
Qy 4092 VTIIIEDVDPQMANLKMELRDALCAIIEEDILTAEGPRIQOAKTSLOBEMDSFSGQ 4151
Qy 2328 ELQPGKVELVSDGLYRKLIISDVHAEDEDYTCADGV-----2366
Db 4152 KVEPITPEVESKYLIISTEESVYFNQSRVKY-LDAPVTVKGVASAVVSDSKQESLKPS 4210
Qy 2367 ----KTSQAQPFVEPQSIITVIRGLQDVTVMEPAPA-----WPECETSI PSVR 2408
Db 4211 EEKESSSESGETEAVATVKIQEAGGLIKEDGPMIHTPLDVTVSEGDIVHLTTSITNAK 4270
Qy 2409 PPKWLLGKTVLQAGNVGLBOEGTVHRLMLRR-----TCSTMTGPVHFTVGKRS 2458
Db 4271 EVNMYFNKLVPSDEKFKCLQDQNTYTLVIDKNVTEHDQGEYVCEALNDS-----GKTAT 4325
Qy 2459 SARL-VVSDIPVLTPLPEKPTGRELOSVLSCDFRAPKA-VQWYKDDTPLSPSEKPKM 2516
Db 4326 SAKLTVVKRAAPVTKRIEPLLEVALGHAKFTCIQSAPNVRQWFKAGRIYSDSKCI 4385
Qy 2517 SLEQMAELRILRLMPADAGVYRCQAGSAHSSTETVTEAREVTVTG-----PLQDA 2567
Db 4386 RSSKYISLELRLQVVDCCGEYTCASNEYGSVCTA-TLTVTPGGEKKYKRLPERKP 4444
Qy 2568 EATEEGWASFCSELSHDEEVEVSLNGMPLVNDSFHIEHGRHRTLVLKSIQADAGIV 2627
Db 4445 EPKEE--VWLKSVLRKRPBEEEPKVEPKL-----EKVKDPAVPEPPPKPVEEVPTV 4497
Qy 2628 RASLSKYSTSARLEVRVKPVVFLKALDLSABEERGTALQCEVSDPE-----AHVV 2678

```
Db 4498 TKRERKIPEPTKVP-EIKPAIPLPAPEPKPKPEAEVKTIKPPPVPEPTPIAAPVTVPVW 4556
Qy 2679 WRKQVOLGSPDKYDFLHTAG-TRGLVVHDVSPDAGLYTCHVGS-----BETRARVRV 2731
Db 4557 GKKAFAK--APKEBAAPKPGIKVGPVKTKTSPTEABERRKLRPGSGGKPKPDPEAPFTYQL 4613
Qy 2732 HDLHVIGITRKLTMVLE-----GESCSPECVLSHESASDPAMWTVGGKTVSGSSSRFOATR 2787
Db 4614 KAVPLKFVKEIKDIIITSEFVGSSAIFECIVSPSTAI--ITWMDKGSNIRESPKHFIA 4671
Qy 2788 QGRKYLIVVREAPSDAGEVFSVR-----GLTSKASLIVRBRPAIHKPLSDQ-WVADGE 2842
Db 4672 DGKORKLHIIDVQLSDAGEYTCVLRLGNKKEKSTAKLVBEELPVRFVKLTBEEVTVVKGQ 4731
Qy 2843 DVELRCELSRAGTPVHMLKDKR-AIRKSQYDVVCEGTMAMLVIRGASLKDAGEYTCBEV 2901
Db 4732 PLYLSCELNKE-RDVWRKDKIVVEKGRIVPGVIGLMRALTINDADDTAGYTVTVVE 4790
Qy 2902 ASK-----STASLHVEEKANCFTTELTNIQVBEKGTAVFTCKTEHPAATVTVWRKGLLELRA 2957
Db 4791 NANNLECSCKVVEVIRDLVKPIRDQHVKPKGTAFACDIAKOTPNIKWFKGYDEIPA 4850
Qy 2958 --SGKHQPSQGLTLRLTISALEKADSDTYTCDICQAQSRQALLVOGRVHIIIEDELDVD 3015
Db 4851 EPNDKTEILRDGNHLYLKIKNAMPEDIAEYAVEIEGKRYPAKLTIGEREVELLXPIDVT 4910
Qy 3016 VOBGSSATFRCKRISPANYPVHWFLDKTPLHANELNEIDAOPGGVHVLTQLALKDQSGT 3075
Db 4911 IYKESASFDAEISEADI-PCQWKLKGBELLRPSPTCEIKAB-GGKRFILTLHKVKLDQAGE 4968
Qy 3076 IYFEAGDQASAAALRVTEKPSVFGRELTDTATITEGEDTLVCETSTCDIPMCWTKDQKTL 3135
Db 4969 VLYQALNAITTAITLVKIEIELDFVPLKDVTVPERRQARPEC-VLTREANVIMSKGPDII 5027
Qy 3136 RGSARCQLSHGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARPVPFQALKDLE 3195
Db 5028 KSSDKFDIIADQKGHILVINDSQFDEGVYTAEVGKKTSARLFTVGIKLFKSPLEBDQ 5087
Qy 3196 VLEGGAAATLRCVLSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVNRLRPDQSGRYS 3255
Db 5088 VKEGETATFVCELSHEKMHVVWFKNDAKLHTSRVTLSISEGKTHLEKWEVTLDDISQIK 5147
Qy 3256 CSPGDQTSATLTVTALPAQFIGKLNRKBEATGATATLRCBSLKTAPVWRKSGSETLRDG 3315
Db 5148 AQVKELSTAQLKVEADPYFTVKLHDKTAVEKDBITLKCEVSKDVPVKWFKDGEIIVPS 5207
Qy 3316 DRYCLRDQAMCELIQIRGLAMVDAEYSCVGEERTSASLIRPMPAHFIRLKHQESIE 3375
Db 5208 PKYSIKADGLRRLIKIKKADLKDKGEYVDCDGTDKTKANVTVEARLIEVEKPLYGVEFV 5267
Qy 3376 GATATLRCELS-----KAAPVWRKGRSIRDGDHSLRQDQAVCELOICGLAVA 3425
Db 5268 GETAHFEIELESDPVHGWKULKQPLTASPCCEIIEDGKH-----ILLHNCQLGM- 5319
Qy 3426 DAGBYSCVCGBERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCELSKVAPV-EWRKG 3484
Db 5320 -TGEVFSFOAANAKSAANLKVKELPLIFITPLSDVKVFEKDEAKPECEVSRBPKTFRWLKG 5378
Qy 3485 PENLRDGRYTLROEGTFCLOICGLAMADAGEVLCVCGQERTSATLTIRALPARFTEDV 3544
Db 5379 TOEITGDDRFELIKDGTKHSVWIKSAAPFEDBAKYNFEABDHTSGTKLIIEGIRULKFLTPL 5438
Qy 3545 KNOBAREGATVLOCEL-NSAAPVWRKGSBTLRDGRYSIRQDGTKCELIQIRGLAMADT 3603
Db 5439 KDVTAKEKESAVFTVELSHDNIRVKWFKNQDRLHTRSVSNQDEBKTHSITFKDLSIDDT 5498
Qy 3604 GEYSCVCGQERTSAMLTVRALPIKFTTEGLRNEEATEGATVLRCELSKM-APVBEWKGHE 3662
Db 5499 SQIRVEAMGMSSEAKLTVLEGDPYFTKGLDYTGVEKDEVTLQCEISKADAPVKWFKDQK 5558
Qy 3663 TLRDGRHSLRDQARCELIQIRGLVAEDAGVLCWCKERTSAMLTVRAMPSKEIEGLRN 3722
Db 5559 EIKFSKNAVIKDGGKRMJLIUKKALKDIDGTYTCCDGTDKTSKGLDIEDREIKLVRPLHS 5618
Qy 3723 EE----- 3724
Db 5619 VEVMETETARFETISEDDIHANWKLKGEALLQTPDCEIKEEGKIHSLVHNCRLDQTOGG 5678
Qy 3725 -----ATBGDTATLWCELS-KAAPVWRKKGHEHTL 3752
Db 5679 VDFQAAANVKSAAHLVRKPRVIGILLRPLKDVTVTAGETATFDCELSYEDIPVEMYLKGGKL 5738
Qy 3753 RDGRHSLRDGSRCELIQIRGLAVVDAGEYSCVCGQERTSATLTIRALPARFIEDVKNQE 3812
Db 5739 EPSDKVVRSEGKVHTLUTLRDVKLEDAAGEVOLTAQDKFTHANLFVKGPPEVFTKPLSDQT 5798
Qy 3813 AREGATVLOCELSKA-APVWRKGSBTLRDGRYSIRQDGTCELOIHLGSLVADTGEYS 3871
Db 5799 VEEGATVLECEVSRENAKVKWFKNGTEILKSKYEIVADGRVRLKVIHDCPTEDIKTYT 5858
Qy 3872 CVCQERTSATLTVRAPQVPRBPQLSLOABEGSTATLQCELSBPTTVVWKSGLQLOA 3931
Db 5859 CDAKDFKTCNLTNVPVPHVBFRLPLTDLQVREKEMARFECELSRENAKVKWFKDGAIBKK 5918
Qy 3932 NGRREPLQCTAELVLQDLQREDTGEYTCCTCSQATSATLTVTAAAPVRLRELQHOEVD 3991
Db 5919 GKTYDIISKGAIRILVINKLLDDAEYSCVTRTARTSGMLTVLEEBEAVFTKNLANIEVS 5978
Qy 3992 EGGTAHLCCELSRAGASVEWRKGSLOLPFCAKYQMVODGAAEILLVRGVEQEDAGDYTC 4051
Db 5979 ETDTIKLVCEVSKPGAIEVWYKGDDEEIIETGRVEILLTEGRKILVIQNAHLEDAAGYNCR 6038
Qy 4052 TGHQTSMAISVVRVPRPKFTRLOSLEQETGDIARLCCQLSDABSAGVOWMLKEGVELHA 4111
Db 6039 LPSRSRTDGKVKHLEAAEFISKQNLSEILSEKAEFVCSIS--KESFPVQWKRDKDKLES 6096
Qy 4112 GPKYEMRSQATBELLIHQLEAKDGTGEYACVTCGQKTAASIRVTEPEVTVIRGLVDAEVT 4171
Db 6097 GDKYDVITADGKRVLVVKDQATLQDMGYTVVWGAARAAHLTVIE-KLRIVVPLKOTRVK 6155
Qy 4172 ADEBVEFSCVSRAGATGVQWCLQGLPLQSNVEVTEVARDGRHTRHLKGVTPBEDAGTVS 4231
Db 6156 EQQEVVFNCEVNTGAKAKWFRNEEAIFDSSKY--IILQKDLVVTIARDAHLDDQANYN 6213
Qy 4232 FHLGNH-----ASSAQLTVRAPEVTVILEPQDVOLSEGQASFOCLRSLRSGQBARWALG 4286
Db 6214 VSLTNRHGENVKSAANLIVEEDLRIVEPLKDIETMEKKSFTWCKVNRNLN-VTLKWTQN 6272
Qy 4287 GVPLQANEMNDITVEQGTLLHLTLHKVTLLEDAGTVSPHVGTCSESAOLK-VTAKNTVVVG 4345
Db 6273 GEEVPPDNRVSVYRDK-YKHMILTIKDCGFPDEGEYIVTAGQDKSVAELLIEAPTEFEVH 6331
Qy 4346 LENVEALEGGEALFECOLSQPEVAAHTWLLDDBPVRTSENAEVVFFPENGRLHLLKRLR 4405
Db 6332 LEDQTVTEFDVAFVSCQLSR-EKANVKWYRNGREIK--EGKYYKFKDGSIHRLI IKDCR 6388
Qy 4406 PQDSCRVTVFLAGDMVTSAFITVGRWRLIELEPLKNAAVRAGAQAARFTCTISEAVPVGEAS 4465
Db 6389 LDDECEYACGVEDRKSRARLFVEIIRPPQDILEAPGADVVFVLAELNK--DKVEVQ 6446
Qy 4466 WYINGAAVQDDSDWTVTADGSHQALLRSQAOPHAGEVTFACRDVAVASAR----- 4516
Db 6447 WLRNMVWVQDGH-QMMSGKIHRLQICDKPRDQGEYRIFIAKDEKARAKLELAAPKI 6505
Qy 4517 ----- 4518
Db 6506 KTADQDLVDVGKPLTWVPYDAPKAEABWFKENEPLSTKTIIDTTAEQTSFRILEAKKG 6565
Qy 4517 -----LTVLGLPDPPEAEVVAHSSHVTWLSWAAPMSDGGGLC 4555
Db 6566 DKGRYKIVLQNKHGKABGFINLKVIDVPGEVRLNLEVTETFDGVSLEAWEBPLTDSKII 6625
Qy 4556 GYRVEVKEGATGOWRLCHELVPGPECVWDGLAPGET-YRPRVAAGVPVGAPEVPHLPQTV 4614
Db 6626 GYVVERDIRKRTWVLATDRAESCEFTVTLQKGGVEYLFVRSARNRNVGTGEPVETDNPV 6685
```

Qy 4615 RLAPPPKVPPOP-----SAPESRQVAGEDVSLELVVAEAGEVITHKGM 4660
Db 6686 E-ARSKYDVPGLPLNVTITDNRFGVSLTWPEYDGGAAETNYVIELRDKTSIRWDTAM 6744
Qy 4661 ERIQGRFVVSQGRQOMLKGFTABDQCEYHCGLAQGSICPAAATFOVALSPASVDE 4720
Db 6745 T-----VRAEDLSATVDVVEGOEYSFVRQAQNRIGVCK-----PSAATPFVKVADPIERP 6795
Qy 4721 APOPSLPEAAQOGLHLLBALARKRMSREPTLDSISELPEEDGRSQRLOPQAEAEVAP 4780
Db 6796 SPPVNLTSDDQTSQSVQLKWEPPKLD--CGSPILGYIIEECB--GKDNWIRCNM-KLVP 6850
Qy 4781 DLSGYSSTADELARTDADLHSTSDDESRAGTSLVLYLKAGRPSTPLASKVGAPAA 4840
Db 6851 ELTVKVTGLEK-----GNKLYRVAENKAGVSDPSEI----- 6883
Qy 4841 PSVRPQQOQELAAVRPPLGDLSTKD-LGDPMSMKAAVK-----IQAAFGYK 4887
Db 6884 -----LGPLTADDAFVEPTMDLSAFKDGLELVVNPITILVPSTGY- 6924
Qy 4888 VRKMKQOEGMFHSTFGDTBAQGDALRLBECVASKADVARMLKDGVELTIDGRHHHID 4947
Db 6925 -----PRPTATWCFGDKLETGDRVKMKTLSA----- 6951
Qy 4948 QLGDGTCSELLTAGLDRADAGCYTCOVSN-----KEGOVTHSA 4984
Db 6952 -----YAELVISPERSDKGIYTLKLENRVKTIISGEIDVNVIAAPSAPKELKFGDITKDS 7006
Qy 4985 CVWVSGSEASESSGGLDDAFRAARLRLPRTK----- 5021
Db 7007 VHLT-----WEPDDGGSPLYGYVVEKREVRKTKWMDFTDLEFTVPDLVQGEYLF 7062
Qy 5022 -----SPAESVDEELFSLADEGPAPPEPADQTYREDSHFI----- 5058
Db 7063 KVCARNKCGEPAYV-DEPVNMSTPATVPDPENVKWRDRANSIFLTWPPKNDGGS 7121
Qy 5059 -----CIR-----FALTEARQAVTRFQB--MFATLGIGVEIKLVEQGRPRVEM 5100
Db 7122 IKGVIVERCRPGSKWVACGPFAETKMEVTGLEEGKYA-----YRVKT 7166
Qy 5101 CISKETPAPVPPPEPLPSLLTSDAAPVLTB--LQOQEVQDGYPVSCVVTGQPMPSVR 5158
Db 7167 LNRQAKSPSPTEIEIAVDTOEABEIFLDVKLLAGLTVKAGTKIELPATVITGKEPKIT 7226
Qy 5159 WFKQGLLEEDHYMINEDQGGHQLIITAVVPADMGVYRCCLAENSMGVSSTKAELRYDL 5218
Db 7227 WTKADMILKQKRITI-ENVPKSTVTIVDSKRSDTGYIIEAVNVCGRATAVVEVNV-L 7284
Qy 5219 TSTDYDTAADATESSSYFSAOGYLSRSREOQTESTTDEGQLPOVVEBLRLQVAPGTRLA 5278
Db 7285 DKGPAPAFDITDVTNESCLLTNPPRDDGGS----- 7316
Qy 5279 KFLQVKGYAPAPRLWFKQDGLPTASAHIRMTGKILHTLEIISVTRSDSGOYAAVISA 5338
Db 7317 -----KITNYVER-----RATDSEVWHKL-----SSTVKTDFKATKLI PN 7353
Qy 5339 MGAAYSSARLLVRGDEPEEPKASDVHQLVPPRMLERFTRPKVKKGS-SITFSYKVE-- 5395
Db 7354 KEYIFRVAENMYGAGEPVQASPIAKYQFDPDPGPPTRLEPSDITKDAVTLTWCEPDDG 7413
Qy 5396 GRPVTWHLREABRGVLWGPDPGTYTVASSAQOHSLLVDVGRHQGTY--TCIASN 5453
Db 7414 GSPI-TGYWV-----ER-----LDPDTDKVRNCNMKPVKDTTVRVKGLTNKKKYRFRVLAEN 7464
Qy 5454 AAGQALCSASLHVSLP-KVBEQEKKEALISTELQ--TTOAISAOGLETASPADLGGQ 5510
Db 7465 LAG-----PKPSKSTEPILIKDIPDPWPPGKPTVKDVKTSVR-----LNWT 7508
Qy 5511 RKEEPLAAKEALGHSLEAVGTEEPQLKTSQITEMVSAKITQAKLVPG-GDSDESKT 5569
Db 7509 KPEHDGAKIESYVIEMLKTGTDEWVR-----VAEGVPTTQHLPLGLMEGQEYSFR 7559
Qy 5570 PSASPRHGRSPSSSIQSSSESEGDGARGEIPDIYVVVTADYLPGLGAEQDAITLREGQV- 5628

Db 7560 VRANKAGESEPSB-----PSDPVLCREKLYP 7586
Qy 5629 -----VEVLDAAH-----PLRWLVTRTKP-----TKSSPSRQGM--VSPAYDLRRLK 5667
Db 7587 PSPRWLEVINITKNTADLKWTVPEKGGSPITWVIVEKRDVRRKGMQVTDITVKTCT 7646
Qy 5668 LSP--ENG-----AAAPFPPEGAUSEDEYKARLSSVIOELLSEQAFVBEQLQSHH 5719
Db 7647 VTPLTEGSLYVFRVAEE-----NAIGOSDY-----TEIEDSVLAKOTFT-----7685
Qy 5720 LQHLERCPHPIAVAGQKAVIFRNVDRDIGRPHSHSFLOELQOCDTDDDDVAMCFI-KNOAAF 5778
Db 7686 -----TPGPPYALA-----VVDTKRHVDLKWEPKPKNDGGRPIQRYVIEKERLG 7730
Qy 5779 EOYLEFLVGRV-----QAESVVVSTAIQFYKYYAE-----BALLAGDPQ 5819
Db 7731 TRWVK--AGKTAGPDCNFRVTDVIEGTEVQ--FQVRAENEAGVGHPSBPTILSIEDPTS 7786
Qy 5820 PPPPLQHYLB-----QOVER-----VQRYQALLKELIRKA 5851
Db 7787 PPSPLDLHVTDAGRKHIAIAWKPPKNGGSPIIIGYHVEMCPVGTCKMVRVNSPIKDLK 7846
Qy 5852 RNRQNCALLEQAYAVVSGALPORAENKLHVS-----LMENYPG-----TLEALGEPIRQG 5900
Db 7847 FKVEGVVPDKEY-----VLRVAVNAICVSEPSISENVVAKDPCKCTID-----LETH 7897
Qy 5901 HFIYWEG-----APGARMWPKHNRHV-----FLFRN-----HLVICKP-RRDS 5938
Db 7898 DIIVIEGKLSIPVPFRAVPVPTVWHKDGKEVKASDRLTMKNDHISAHLEVPKSVRADA 7957
Qy 5939 RTDVTSVYFRNMKLSIDLNDQVEG-----DRAFEV-----WQ----- 5973
Db 7958 GIYTIIT--LENKLGASATASINVKVLGPGCKDIKASDITKSSCKLTWEPPEFDGGTPI 8015
Qy 5974 -----EREDSVKYLLOARTAIKSSW-VKEIC-----GIQORLALPV- 6010
Db 8016 HYVLEREAGARTYIPVMSGENKLSWTVKDIPNGEYFFRVKAVNKVGGGEYIELKNPVI 8075
Qy 6011 -----WRPPD--FEEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDPHILIE 6064
Db 8076 AQDPKQPPDPVDEVHNPTAE-----AMTITWKPPL--YDGGSKIM--GYIEBK 8121
Qy 6065 DPDGS-----CALILDSLTCVDSG-----QWCFPAASAGNCSTLGKILVQVPP-RF 6110
Db 8122 IAKGEERWKRCNEHLVPIITYTAKLEGEKEYQFRAENAAG-----ISEPSRATPPTKA 8177
Qy 6111 NVKVRASPFV-----EGEDAQFTCTIECAPYQIRWYKD-----CALLTT 6150
Db 8178 VDPIDAPKVILRTSLEVRGDEIALDASISGSPITITWIKDENVIVPEEIKKRAAPLVR 8237
Qy 6151 GNKFO-----TLSEPRSGLLVLVIRAASKEDLGLYECBLVNLRSARASA 6195
Db 8238 RRKEVQEEEPFVPLTQRLSIDNSKKGESQLRVDRSLRDPDGLYMIKVENDHGIKAPC 8297
Qy 6196 ELRTQS-----PMLQAQEOCHREQLVAVE-----DTTLERAOQE-----VT 6232
Db 8298 TVSLDTPGPPINFVEDIRKTSVLCKWEPLPDGGSSEIINYTLEKKDKTKPDSEWIVVT 8357
Qy 6233 SVLK-----RLLGPK-----APGS-----TGLDTGPGPCP----- 6258
Db 8358 STLRHCKYSVTKLIEGKEYLFRVAENRFGPGPCVSKPLVAKDPFGPPDAPDKPIVEDV 8417
Qy 6259 -----RGAPAL----- 6264
Db 8418 TSNSMLVKMNEPKONGSPILGYWLEKREVNSTHMSVNSLLNALKANVDGLLEGLTVYF 8477
Qy 6265 ---QETGQPPVTGTSEAPVPPVPQPLLHEGEQEPEEATARAQEWTVPTIRMEGAAPG 6321
Db 8478 RVCAENAAGP---GKFSPPSPDKTAHDPISPGP--PIPRVTDTSSTSTIELEWEPAPNG 8532
Qy 6322 AG-----TCCELLWD-----VHSHVVRETTQ-RTYTYQAIDHTTA----- 6354

Db 8533 GGEIVGVYFVDKLVGNTKNSRCTEKMIKVRQVTVVKEIREGADYKURVAVNAAGPPGE 8592
Qy 6355 -----RPPSMQVETIE-----DYQAQTGGTAQFAEIIIEGDPQSPVWYKDSVOLVDST 6401
Db 8593 TQPVTVAEQPEPPAVELDVSKGGIIMAGKTLRIPAVVTGPPVTKVWTKEEGEL-DKD 8651
Qy 6402 RLSQOQEGTYSVLVRHVASDKAGVYTCLAQNTGGQVLCARLLVLGGDNPEP-----DSEK 6457
Db 8652 RVIDNVGTGKSLIIKDALRKDHGRYVITATNSCGSKFAAARVEVF-----DVGPPVLDLKP 8708
Qy 6458 QSHREK--LHSPYEVKEEIGRVFGVRVQKHGKNILCAAKFIPLRSTRQAQVREDI 6515
Db 8709 VVTRAKMLNDSWDEGGSEITGFI--IERKDAKM----- 8743
Qy 6516 LAALSHPLVTGLDQFETRKTLILILELCSSEELLDRLYRGVVTAEVVKVYIQOLVREL 6575
Db 8744 -----HTWRQPIETERS-KCDITCLLSGQ 8766
Qy 6576 HYLHSHGVL-----HLDIKPSNIMLVHPAREDIKICDFGFAQNTPAELOFSQVGS 6626
Db 8767 EY--KFRVIKAKFKGCGPVEIGP--ILAVDP-----LGP 8797
Qy 6627 PEFTVSPETIIQNPVSEASDIWAMGVISYLSLTCCSPFAGESDRATL-LNVLEGRVSMSSP 6685
Db 8798 P--TSPE-----RLTYT-----ERQRTITLWKEPRNSGGSP 8828
Qy 6686 MAALHSE-----DAKDFIKATLQAPQAPRPSAAQCLSHFPWFLKSPAEBAHFINTKQKFL 6741
Db 8829 IQGYIIEKRHRDKPDF-----ERVNKLCPPTTSFLVENLDEHQMYEVRK----- 8873
Qy 6742 LARSWQRSLMSYKSLIWNRSIPELLRGPPSPSLGVARHLCRDTGGSSSSSSSSDNELA 6801
Db 8874 -----AVNEIGESEPSLP-LNVV-----IQDDEVP 8897
Qy 6802 PPARAK-----SLP--PSPVTHSPL-----LHPRGFLRPSASL-----PREAE 6837
Db 8898 PTIKLRLSVRGDTIKVAGEPVHVPADVTGLPMPKIEWSKNETVIERGTDLAQITKGEVS 8957
Qy 6838 ASERSTEAPAPASPEGAPPAACQVPRHSVIRSLFYHQGESPEHGALAPGRRRHPAR 6897
Db 8958 RSEAKTELSIPKAVREDKGTVTVNASNLGVSFRNVHVVDYR-----PSPPRNLAV 9009
Qy 6898 RRHLKGGYIAGALP--GLREPLMEHRLVLEEAREEOA-----TLAK-----A 6940
Db 9010 TDIKAESCYLTDWAPLDNGSGE--ITHYVIDKRDASRKKAEBEVTNTAVEKRYGIWKLI 9067
Qy 6941 PSFETALRLPASGTHLAPGHSLSLEHDSPTSPRPSSEACGEAQLPSAPSGAPIRDMGH 7000
Db 9068 PNGQYEFVRVAVNKY--GISECKSDKVVIQDP-----YRLPGPP--GKP-KVLAR 9113
Qy 7001 PQSGKQLPST-----GHPGTAQ--PERPSPDP--WGQAPAPCHPKQG-----SAPO--E 7045
Db 9114 TKGMLVSWTTPPLDNGSGSPITGYWLEKREESPYWSRVSRAPITKVGLKGVEFNVPRLL 9173
Qy 7046 GCS-PHPAVACPPGSPPGSKCAPLVPSP-FLGQOAPAPAKASPPPLDSXMGPGDI 7103
Db 9174 GVKYQFRAMAINAAGIGPSE-----PSDPEVAGDPFIPPPG-PSCEPVKDKT-KSSI 9224
Qy 7104 SLPRGPRPGPCSPGASASQSSSVSLRVGSSOVGTGPGSLDAEGW--TQEAEDLSDS 7161
Db 9225 SLGKWP--PAKDG-----SPIKGYIVEMQEGT-----TWKRVNEPDKLIT 9266
Qy 7162 ----TPTLQRPQEQATWRKFLSGRGGYAGVAGYGTTFAGDAGGMLGQGPWARIAMAV 7217
Db 9267 CECVVPNLKE-----LRKYRF-----RVK-AV 9287
Qy 7218 SQSEEEQEEARBSQSEEQEAEASPLPOVSARPVPEVG-----RATRS 7264
Db 9288 NEAGESPFTGTGIPATDIOEE-----PEVFT-----DIGAODCLVCKAGSQIRIPAVI 9337
Qy 7265 SPEPTP---WEDIQVSLVQ---IRDLSGAEEAATISLDISEVDPAVLNLSLDYDI--K 7316
Db 9338 KGRFTPKSSWEFDGKAKKAMKGDVHDIPEDAQLTAENSSVIIPECKRSHTKYSITAK 9397

Qy 7317 YLPPEFMIFRKVPKSAQPEPPSPMAEBELAEFP--EPTWMPFGEIG----- 7360
Db 9398 NKAQKQKANCVRKVMQDVPVGGPKDLKVSIDIIRGSCRLSKWMPDDDDGGDRIKGVIEKRTID 9457
Qy 7361 -----PHAGL-----EITESESDVALLABAAVGRKRKWSPPSRSLPHFFGRHLPL 7406
Db 9458 GKAWTKVNPDCGSTTFVVPDILLSQQYFFRVRARENFGIPPVETIQRTTARDP--IYPP 9515
Qy 7407 DEPAELGLRERVKASV-----BHSIRLKGPRPGLSEKGE----- 7440
Db 9516 DPPIKLKILGITKNTVHLSWKPKKNDGSGSPVTHYIVELAWDPTGTKKBAWRQCNKRDVE 9575
Qy 7441 -----PPRKPGGLASFRLSGLKSW-----DRAPTF-LKELSDETIVL 7476
Db 9576 ELQFTVEDLVEGGBYEFRKAVNAAGVSKPSATVGPCQCDQCPDMPSPSIDLKEFNE--VEE 9633
Qy 7477 GQSVTLACQVSAQAPAAQATWSKOGAPLESSRVLISATLKNFOLL--TILVVVAE-----D 7530
Db 9634 GTNNVIVAKIKGVPPFTLTWFKAPPKPDNKEPVLVYDTHVKNKLVDDTCTLVIPOQSRSD 9693
Qy 7531 LGVVTCSVNAIGTVTTTGVLRKAERPSSPCPDIG-----EYVADGVLLWVKPVEYGP 7585
Db 9694 TGLTITVANLGTASKEMLNLVREG-----PPVGPIKFESVADOMTLSWFPKDDGG 9749
Qy 7586 ---VTVIIVQ-CSLEGGSWTTLASDIFDCCYLTSKLSRGGTYTTFRTACVSKAGMG-PYSSP 7640
Db 9750 SKITNVYIEKREANRKTWVHVSSEPKECTYIPKLLGHEVVFPRIMAQNKYVIGEPDSE 9809
Qy 7641 SE 7642
Db 9810 PE 9811
RESULT 10
Q8WZ42
ID Q8WZ42 PRELIMINARY; PRT; 34350 AA.
AC Q8WZ42;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Titin.
GN Name=TTN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309627; PubMed=10850961;
RA Freiburg A., Trombitas K., Hall W., Cazorla O., Fougereousse F.,
RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,
RA Granzier H., Labeit S.;
RT "Series of exon-skipping events in the elastic spring region of titin
RT as the structural basis for myofibrillar elastic diversity.";
RL Circ. Res. 86:1114-1121(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21573839; PubMed=11717165;
RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,
RA McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,
RA Labeit S.;
RT "The complete gene sequence of titin, expression of an unusual -700
RT kDa titin isoform and its interaction with obscurin identify a novel
RT Z-line to I-band linking system.";
RL Circ. Res. 89:1065-1072(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Centner S.B.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277892; CAD12456.1; -.
DR PDB; 1TNM; NMR; @=33480-33579.
DR PDB; 1TNN; NMR; @=33480-33579.

DR Genew; HGNC:12403; TTN.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004896; F:hematopoietin/interferon-class (D200-domain...); IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0008307; F:structural constituent of muscle; IEA.
DR GO; GO:0007517; P:muscle development; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR Pfam; PF00041; fn3; 132.
DR Pfam; PF00047; ig; 3.
DR Pfam; PF00069; PKinase; 1.
DR Pfam; PF02818; PPAK; 53.
DR ProDom; PD000001; Prot.kinase; 1.
DR SMART; SM00060; FN3; 132.
DR SMART; SM00408; IGC2; 65.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00933; FGXY KINASES_1; UNKNOWN_1.
DR PROSITE; PS00853; FN3; 132.
DR PROSITE; PS00835; IG LIKE; 143.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; UNKNOWN_1.
DR SEQUENCE 34350 AA; 3816189 MW; 5B1120058A7CE58A CRC64;

Query Match 8.3%; Score 3427; DB 2; Length 34350;
Best Local Similarity 21.5%; Pred. No. 9.8e-101;
Matches 1839; Conservative 1183; Mismatches 3528; Indels 2005; Gaps 336;
QY 10 PRFTTRKAFVYVSGKDATLSQIVGNPTPOVSWKDDQPVTAGARFRLAQQDGLYRLTI 69
DB 3239 PQVQLQPVTVQSGKPARFCNAVISGRPOKLISWYKEQLLSTGPKCFPLHDGQBYTLL 3298
QY 70 LDALGDSQGVYCRARNAIGFAFVAVGLQVDA-EAACAEQA-----PHFLRLPTSIRV-RE 123
DB 3299 IEAPPEDAAVTCFAKNDYGVATTSSASLSVEVPSVDPQEMVYPPAIITPLQDTVTSE 3358
QY 124 GSEATFCRCVGSPPAVSWKDGRRLGEPDGRVPRVVEELGEASALRIARAARPRGGTYE 183
DB 3359 GQAPRFQCRVSGTDL-KVSWYSKDKKI--KPSRFERMTQFEDTYQLEIAEAYPEDEGTYT 3415
QY 184 VRAENPLCAASAAALVVD-----SDAADTASRPGTSTAALLAH 222
DB 3416 FVSAVAVQVSTANLSLEAPESILHERIEQBIEMEMKEFSSFLSABEGLHSAEL--Q 3473
QY 223 LQRRREARAGCAPSP-----PSTG-----TRTCTVTBGRHARLSYVTVGERPK 266
DB 3474 LSKINETL--ELLSESPVYTKFDSEKEGTGPIFIKEVSNADISMGDVATLSVTVIGIPK 3531
QY 267 PETVWKDQGLVTEGRRHVVEDAQENFVKILPKQSDRGLYTCTASNLVGGQTVSSVLV 326
DB 3532 PKIQWFFNGVLLTPSADYKFVFDGDDHSLI-ILFTKLEDEGEYTCMASNDYKTKTCSAYL 3590
QY 327 VVRE-----PAPV-FKKRLQDLEVREKESATF-LCEVPQPS 361
DB 3591 KINSGEGHKTETESAVAKSLEKLGKGCPPHFLKELPIRCAQGLPAIFETVVVGEAP 3650
QY 362 EAAMFKEETRLWASAKYGI-----EBEGTERLLTVRNVGADDDAVYICR----- 405
DB 3651 TVTWFKENKQLCTSVYITIHPNNGSGT--FIVNDPQREDSGLYICKAENMLGSECTCA 3707
QY 406 -----TPEGSRVVAELAVQGNLRLKL--PKRTAVRGDGTAMFCV- 442
DB 3708 ELLVLEDTMDTPTCKAKSTPEAPDFPPLKGPFAVEALDSEQEIATFVKDITLKAAAL 3767
QY 443 -----ELA--VPVG--PVHWRNQEEVAVAGRVVAISAEGRHRLTISQ 481
DB 3768 ITEENQQLSYHIAKANELSSQLPQAQELQSILEQDK-----LTPESTREFLCIN- 3818
QY 482 CCLEDVGQVAF-----MAGDCQSTSTFCVSAAPKPPQLQPPVDPVK-----A 523
DB 3819 -----GSIHQPLKEPSNPQLQIQVQSKQTFSEKILMPPEPTQAVLSDEKIFPSAM 3872

QY 524 RMESSVILSNP-----PPHGERPVT-----IDGYLVEKKKLGTYTWTIRCHAEAWATP 572
DB 3873 SIEQINSUTVPLKTLTLLAEPEGNYPQSSIEPPMHSYLT-----VABEVLSPP 3919
QY 573 -ELTVADVAEENFQFRVSALNSFQSPVLEFPFGTVHAPKLA-----VRTPLKAVQAV 625
DB 3920 KEKTVSDT-----NREQRTVLQKEAQSALI-----LSQSLAEGHVESLQSPDVWISQV 3968
QY 626 -----EGGEVTVFVDLTVASAGE-----WFLDQAL-----KASSVYEIH 660
DB 3969 NYEPLVPSEHSCTEGKILIESANPLENAGQDSAVRIEKGSLRPLALEEKQVLLKEH 4028
QY 661 CDRT-----RHTLTIREVPASLHGQAL-----KRVANGI--ESSIRMEVRAAPGLT 704
DB 4029 SDNVVMPDPPIETESKREPVAIKKV-QEVQGRDLLSKESLLSGIPSEQRLNLKIQICRALQ 4087
QY 705 ANKPEAAAAAREVLARLHEEAOLLAELSQAATAVTLKDGRTLSPGPKVEVQA----- 756
DB 4088 A-----AVASEQPGLFSE-----WLRNIE-----KVEVEAVNITQEPBR 4120
QY 757 -----SAGRRV-----LLVRDVARDAGL-----YE-----CVSRGGRYAY 787
DB 4121 HIMCMYLVTSAKSVTEETVIIIEDVDPQMANLKMELRDALCAIIYEEIDILTAEPRIQ- 4179
QY 788 QLSVQGLARFLHKDAGSCVDVAGCPAQFECETSEAHVHVHWHYKDGHELGHSGERFLOE 847
DB 4180 -----QAKTSLQEBM-----DSFGS--SOKVEPITEPEVESKYLISPEVSFY- 4221
QY 848 DVGTRHRLVAAT-VTRQDEGTVSCRVGDSVDRLRVSEPKVVFPAK-----EOLARRKLQ- 901
DB 4222 NVQSEVKYLDATPVTK-----GVASAVVSEKODESLKPEEKEESESSESTEEVATVKIOE 4278
QY 902 -----ABAGASATLSCEVAQAQTEVTWYKDGKLLSSSKV-CME 939
DB 4279 AEGGFIKEDGPMIHTPLVDTVSEGDIVHLTTSITNAK-EVNWYPENKLVPSDEKFKCIQ 4337
QY 940 ATGCTRLRVVQAGQAD-AGEYSCEA-----GGQRLSFLHDYKEPKVFAKQDVAHSEVQA 994
DB 4338 DON-TYTLIDKVNTEHQGEYVCEALNDSKGTATSAKLVVVKRAAPVIRKIEPLEV-- 4394
QY 995 EAGANATLSCEVAQA-QAEVVMWYKDGKLLSSKLVHVEAKGCRRLRVVQAGKTDAGDYS 1053
DB 4395 ALGHLAKFTCHIQSAPNVRPWFQAGRIEYSDKCSIRSSKYISLEILRTQVVDGCEYT 4454
QY 1054 CEARGQVSP-----RLHITE--PKMFAKESVHNEVQAEAGASAMLSCEVAQAQ- 1106
DB 4455 CKASNEVGSVCTATLTVEAVPTFLSRPKSLTTFV---GKAAKFTCTVTGTPIETI 4510
QY 1107 WYKDGKLLSSSKVGMVEVKGCTRLRVLPQAGKADAGEYSCA-----GQORVSFHLHITEP 1162
DB 4511 WQKDGALSPSPNWKISDAENKHILELSNLTIDRQGVYSCASKNKFGADICQAEILIID- 4569
QY 1163 KGVFAKESVHNEVQAEAGATTLMSCEVAQ-OTEVTWYKDGKLL--SSSSKVRMEVKG 1219
DB 4570 KPHFIKEL---EPQVSAINKKVHLECEQVDEBRKVTVTWSKQKQLPPGKDYKICFEDKTA 4626
QY 1220 TRLRVVQGVKADAGEYSC-----EAGQORVSFOLHITEPKAVFAKEQLVHNEVTRTEAGAS 1275
DB 4627 TLEIPLAKL--KDSGTVYCTASNEAGSSSCSATVTVREPPS-FVKK--VDPSSYMLPGES 4681
QY 1276 ATLSCEVAQAQ-ETWYKDGKLLSSSKVRIEAGCMRQLVVOQAGQADAGEYTCFA-- 1332
DB 4682 ARHLCKLKGSPVIQVTFWPKNNKELSESNTVMYFVNSEAILDITDVKVEDSGSYSCAVN 4741
QY 1333 --GGQRLSFHLDVSEPKAVFAKEQLAHRKVQAEAGAIATLSCEVA-QAOTEVTWYKDGK 1389
DB 4742 DVGSDSCSTETIVIEKPPSFIKLEPA-----DIVRGTNALLQCEVSGTGFPELSWFKDKKQ 4797
QY 1390 LSSSSKVRMEAVGCTTRLRVVQAGQADTGEYSC-----BAGQORLSFSLDVAEPKVVFAKE 1445
DB 4798 IRSSKKYRLFPQKLSVLCLEIFSFNSADVGECVAVNEVGKCGCWATHLLKEPPTFVKV 4857
QY 1446 QPVHREVOAQAAGSTTSLCEVAQAQ-TEVMWYK-----DGK-KLSFSSKVRMEAVGCT 1496

Db 4858 D-----DLIALGGQTVTLQAAVRGSEPI SVTWKQGBVIREDDGKIMSPFNGVAV----- 4907
Qy 1497 RELVVOQAQADAGEYSC-----EAGSORLSFHLHVAEPKAVFAKEQBPASREVQABAGTSA 1552
Db 4908 --LIIPDVQISFGGKYTKLAENAGSGQTSVGLIVKEPAKIIER-----AELIQVTAGDPA 4961
Qy 1553 TLSCEVA-QAOTEVTVYKDGKLLSSSSKVRMEAVGCTRLVLVQEAQADAGEY---SCK 1607
Db 4962 TLEYTVAGTPELKPWKYKDGRLPVLASKYRISEFKNVNAQLKPYSAELHDSGGQTYTEISNE 5021
Qy 1608 AGDQRLSFHLHVAEPKAVFAKEQBPASREVQABAGASATLSCEVA-QAOTEVTVYKDGKLL 1666
Db 5022 VGSSSCETTFTVLDRDIAPEFTKPL-RNVDVSVNGTCRLDCKIAGSLPMRVSMFWDGKEI 5080
Qy 1667 SSSSKVRVEAVGCTRLVLVQQAQADAGEYSCA-----CGQRLSFLHVAELPEQISERP 1722
Db 5081 AASDRYRIAFVGTASLEIIRVDMNDAGNFTCRATNSVSGKSSGALIVQE-PPSFVTKP 5139
Qy 1723 CRREPLVVKHEHEDIILTATLATPSAATVTLKDGVEIRRSKRHETASQGDTHTLTVHGAO 1782
Db 5140 GSKD---VLPGSAVCLKSTFGQSTPLTIRWFKGNKELVSGGSCYITKEALESLLYLK 5196
Qy 1783 VLDSAIYSCRVG--AEGQDFPVQ--VEEVAAPKRLLEPVCGBL---GGTVTLACELS-P 1834
Db 5197 TSDSGTYTCKVNVAGGVECSANLKVKE-PATFVEKLEP--SOLLKKGDATQLACKVTGT 5253
Qy 1835 ACAEVVWRCGNTQPRVGRFOM--VAEGPVRSILVGLRAEDAGEVVCESRD---DHTS 1888
Db 5254 PIKITWAFNDREIIESSKHRMSFVESTAVLRLTDVGI--EDSGYMCBAQNEAGSDHCS 5311
Qy 1889 AOLTV-SVPRVVVKFMSGIST-----VVAEGGEATFOCVVSPSDVAVWVFRDQALLQP 1940
Db 5312 SIVIVKESYFTKEPKPIEVLKEYDVMLLAEVAGTFPPE-----ITWFKONTILRS 5362
Qy 1941 SKFAISOGSAGHSITISDLVEDAQ-----ITVEAGASSAALURVEAPVLFKXKLEP 1996
Db 5363 GRKYKTFIQDLHVLVSQILKFVAADAGEYQCRVTNEVGSSICARSVTLREPP-SFIKKIES 5421
Qy 1997 QTVEBSSSVTLEVELTRPWP-ELRWRNATALAPGNVEIHAEAGARHLVLHNVGFADRG 2055
Db 5422 TSLRGGTAAFOATLKGSLPITVTLWKDSDEITEDDNRMTFENNVAVLSGIEVKKHDG 5481
Qy 2056 PFGCETPDD-----KTQAKLTVEMRQVRLVGLQAVEAREQGTATMEVQLSHA-DVDSGWT 2110
Db 5482 KYVCAKNDAGIQRCSALLSVK-EPATITEEAVSIDVTQGDPAITLOVKFSGTKEITAKWF 5540
Qy 2111 RDGLRPOQPTCHLAVRGPMTLTLUGLRPBDPSGLMVPKABG-----VHTSARLVVTEL-- 2164
Db 5541 KDGQBLTLGSKYKISVTDVTVSILKIISTEKKDSGEYTFEVQNDVGRSSCKARINVLDLII 5600
Qy 2165 PVSFRPRLQDVVTTEKEKVTLCECL--SRPNVDVRLWKDGVLELRAGKTMATAAQACRSL 2222
Db 5601 PPSFTKKLKMDSIKGSDIDLECI VAGSHP-ISIQWFKDDQISASEKYKFSFHDNTAFL 5659
Qy 2223 TIYRCEFAQGVVYVCDHAQDASSASVKVQGRYTYTLIYRRVLAEDAGEIQFVAENAESRAQ 2282
Db 5660 EISQLEGDSGTYTCSATNK-----AGHNQCSGH 5688
Qy 2283 LRVKELPVLVRLPRLDKTAMEKHGVLRCQVSRASAOVRWFKGSOELOQPKYELVSDGL 2342
Db 5689 LTVKEPPYFVEKPOQSDPNPNTRVOLKALVGTAPEMTIKWFKDNKELHSGAARSWKDDT 5748
Qy 2343 YRKLIISDVHAEDEDTYTC---DAGDVKTSAOFFVEB-----QSITIVRGLQDVTV 2390
Db 5749 STSLFLFAAKATDSGTIYICQISNDVGTATSKATLFFVKBPQFIKPSVPLVLRNGQSTT- 5807
Qy 2391 MEPAPAFECE--TSIPSRVPPKWLKG--TVLQAGG-----NVGLEQEGTWH 2434
Db 5808 -----PECQITGTPKIRVSWYLDGNEITAIQKHGISFIDGLATFIQSGARVENSQTY- 5859
Qy 2435 RLMLRRTCSMTGPHVHTVGVKSRSSARLAVVSDIPVVLTRPLEPKTGRLOQSVLSCDPRP 2494

Db 5860 -----VCEARN-----DAGTASCSIELKVKEPPTFI-RELKPVBEVVKVSDVELECEVTG 5907
Qy 2495 APK-AVQMYKDDTPLSPSEKFKMSLEGQMAELRIILMPADAGVYRC-----QAGSAHST 2549
Db 5908 TTPPEVTVLKNRREIRSSKKYTLTDRVSVFNLHTKCDPSDTGEYQOCIVSNEGSCSCST 5967
Qy 2550 EVTVE-----AREVTVTGLQDAEATBEGMASFSCSELSHEDBEVWSLNGMPLYNDSFHE 2604
Db 5968 RVALKEPPSFYKIKENTTVLTKSSATFOSTVAGPPPI-----ITWLKDDQLDEDDNNY 6022
Qy 2605 ISHKRRHTLVLSIQIADAGIVRASSLKVSTSRARLEVRV-----RPVVFRLKALD 2654
Db 6023 ISFVDSVATLQIRSVDNHSG-----RYTCAQKNESGVERCYAFLVLVQEPAQIVERAK 6075
Qy 2655 DLSAERGTALQCEVS-DPEAHVVRKDGQVQLGSPDKYDFLHTAGTGLVHVHDVSPEDA 2713
Db 6076 SVDTEKDPMTLECVAGTPELKVWLKDGQKQIVPSRYFMSFENNVAASFRIQSMKQDS 6135
Qy 2714 GLYXCHV-----GSEBTRARVRVHDLHV--GITTELKTMVELEGBSCSPECVLSHESADP 2767
Db 6136 QOYTFKVENDFGSSSCDAYLAVLQONTIPPSFTKKLTWMDKVLGSSIHMECKVS-GSLPIS 6194
Qy 2768 AMWTVGGTGVSSSRFOATROGRKYILVREAAPSDADAGEVVFVRGLTS-----KASLIVR 2823
Db 6195 AQWFKDGKKEISTAKYRLVCHERSVSVLEVNLELEDTANYTCKVSNVAGDDACSGILTVK 6254
Qy 2824 ERPAALIKPLEDQVAGCEDEVLCELSRAGTP---VHWLKDRAIKRSQKYDVVCEGTM 2880
Db 6255 EPPSFLVKPGRQAI-PDSTVEFKAILK--GTPPFKIKWFKDDVELVSGPKCFIGLEGST 6311
Qy 2881 AMLVIRGASLXDAGEYTC-----EVEASKSTASLHVEBKANCFTBELTNLQVBEKG-TAVF 2935
Db 6312 SFLNLYSDASKTOYTCHVTNDVSDSCITMILLVTPPK-FVKKLEASKIVKAGDSSL 6370
Qy 2936 TCK-TEHPAATVTVWRKGLLELRASGKHQPSQEGUTLRLTISALBAKASDVTYCD-----IG 2990
Db 6371 ECKIAGSPEIRVWFRNEHELPAADKYRMTFIDSAVIMQNNLSTEDSGDFICEAQNPAG 6430
Qy 2991 QAQSRQALLVQGRVH-----IIEDLEDVQVQESSATFRCRISPANVEPVHWFIDKTPL 3045
Db 6431 STSGCTKVIIVKEPPVFSFPPIVETLKNAEV-----SLECELSGTPPFVWVYKDKRQL 6484
Qy 3046 HANELBIDAQPGGYHV-LTLRLQALKDSGTIYPEAGDQDASAAALRVTEK---PSVFSRE 3101
Db 6485 RSSKKYKIASK--NFHTSIHLNVDTSDIGEHYCKAQNEVSGSDTCVCTVKLKEPPRFVSK 6542
Qy 3102 LTDATITEGEDLTUVCTSTCD-IPMCWTQD-GKTLRGSARCQLSHGRHQAQLITGATL 3159
Db 6543 LNSLTVVAGEPAELQASIEGAQPIFVQWLKEKEEVIRESENIRITFVENVATLOFAKAP 6602
Qy 3160 QDSGRYKCEA--GGACSSSIVRVHARPVRFOEALKDLEVLGGAATLRC-VLSSVAAPV 3215
Db 6603 ANAGYICQIKNDGGMENMATLMVLEPAVIVEKAGPMTVTVTCTLECKVAGTPELSV 6662
Qy 3216 KWYCGNNVLPGDKYLSURQEGAMLELVRLNLRPODSGRYSC-----SFGDQTTSATLTVT- 3270
Db 6663 EWKYKGLLTSSQKHKFSFYNNKISSRLSVRQDAGTYTFVQVNNVSKSSCTAVVDVSD 6722
Qy 3271 -ALPAQFIGLKRKEATEGATATLCELSTAP--VEWRKGSETLRDGRDYCLRDQAMC 3327
Db 6723 RAVEPSPFRLKNTGGVGLGASCILECKVAGSSPISAVMFHEKTKIVSGAKYQTTTFSDNV 6782
Qy 3328 ELQIRGLAMVDAABYSC-----VCGEERTSASLITRPMFAHFIRGLRHOESTEGATATLRC 3383
Db 6783 TLQNSLSDSDMGNTCTVAANVAGSDCRAVLTVQEPSP-FVKEPELVLPGKNTVFTS 6841
Qy 3384 ELSKAAP--VEWRKGRSRLRDRHSRLRQDGA VCELOICGLAVADAGEYSCVCGEERTSA 3441
Db 6842 VIRGTPPKVWFRGARELVKDRCNIFYEDTVAEELFNIDISQSSEYTCVWSNNAGQA 6901
Qy 3442 TLTVKAL--PAKFTTEGLRNEEAVEGATMLWCELSKVAP--VEWRKGPENLRDGRYIL 3496
Db 6902 SCTRLFVKPEAPFLKRLSDHSVEPGKSIILESTYTGTLPISVTWKKGDNFTTSEKNI 6961

Qy	3497	QBOETRCELOICGLAMADAGEYLC-----VCGQERTSATLTIIRALPARFIEDVKQ	3547
Db	6962	VTEKTKTILEILNSTKRDAGQYSCIEINEAGRDVCG-----ALVSTLEPPYFVTELEPL	7015
Qy	3548	EAREGATVALQCELSAAP---VEWRKGSSETLRGDGRYSLRQDGTCKELQIRGLAMADTG	3604
Db	7016	EAAVGDSVSLQCQV-AGTPEITVSWYKGDYTKLRTPPEYRYFYNNVATLVFNKVNINDSG	7074
Qy	3605	EYSCVC-----GOERTSAMLTVRALPIKETEGLRNEEATEGATAVLRCELSKMAPE--	3656
Db	7075	EYTCKAENSIGTASSKTVFRIQERQLPPSPARQLKDIEQTVGLPVTLTCLNGSAPIQVC	7134
Qy	3657	WKKGHETLRDGRHSRLRQDARGCELOTRGLVAEDAGEYLCMC-----GKERTSAMLTVRAM	3712
Db	7135	WYRDGVLRLDDENLQTSFVDNVATLKLQTLDSHGQYSCSASNPLGTATSSARLTARE-	7193
Qy	3713	PSK---FIEGLRNEEATGDTATLWCELSKAAP--VEWRKGHETLRDGRHSRLRQDGRSRC	3767
Db	7194	PKSPFPDIPKPSIDVLAGEADPECHVTGAQPMWRTWSKDKKEIRPGGNYTTCVGNTP	7253
Qy	3768	ELQIRGLAUVDAAGYSVCV-----GQERTSATLTVRALPARPIEDVK-NQAREGATAVLQ	3822
Db	7254	HLRLKVGKDSGGYTCQATNDVGKDMCSAQLSVKE-PPRFVKKLEASKVAKQGESIQLE	7312
Qy	3823	CELSKA--APVEWRKGSSETLRGDGRYSLRQDGTCELOIHGLSVADTGEYSVCV-----GQ	3876
Db	7313	CKISGSPEIKVSWFRNDSLHESWKYMNFSINSVALLTINEASAEDSGDYICEAHNGVGD	7372
Qy	3877	ERTSATLTVRAPOVPFREPLOSLQABEGSTATLQCELS-EPTATVWVSKGGLQLOANGRR	3935
Db	7373	ASCSTATLVKAP-PVFTQKSPVCGALKGSDVILQCEISGTPPEPVEVWVKDKQRVSKKF	7431
Qy	3936	EPRLOGCTAELVQDLQREDTGYETTCGSAQTSATLTVTA---APVRFLURELHQHEVDE	3992
Db	7432	KITSKHFDTSLHILNLEASDVGEYHCKATNEVSGSDTCSVKFKEPPRFVKKLSDTSTLI	7491
Qy	3993	GGTAHLCCELSRA-----GASVEWRKGSILQLPFCAKYQMVQDGA-----AELLV	4037
Db	7492	GDAVEL-----RAIVEGFQPIVVWLKD-----RGEVIRESENTRISFDINIA TLQL	7538
Qy	4038	RGVEQEDAGDYTC---DTGHTOSMASLSVRPRPKFKRLQSLQEQTGDIARLCCOLSD	4093
Db	7539	GSPEASNSGYICOIKNDAGMRCSAVLTLEP-ARIIKEPEMTVTGPPFALECVVTC	7597
Qy	4094	AESGAVVQWLKEGVELHAGPKYEMRSGGATRELLIHOLEAKDTGEYAC-----VTGQOKTA	4149
Db	7598	TPE-LSAKWPKDGRSLSDSKHHITFTNKVASLKIPCAEMSDKGLYSFEVKNSVGKSNCT	7656
Qy	4150	ASLRVTEPEV--TIVRGLVDAEVTADBEDVEFSCEVSPAGATGVQWCLQGLPLQSNVTEV	4207
Db	7657	VSVHVSRIIVPPSPIRKLKDNAILGASVVLRECVSGSAPISVGFWDGNEIVSGPKCQS	7716
Qy	4208	AVRDGRITHLRKGVTPEDAGTVSFHLGNHAS---SAQLTVRAPEVTILLEPLQDVQLSE	4263
Db	7717	SFSE-NVCTNLNLSLLESDDTGIYTCVAANAVGSDCSAVLTQPEPSFETP--DSVEVLP	7774
Qy	4264	QODASFOCRLSRASGQBARWALGVPIQANEMDITVEQGTLLHLLHKVTTLEDAG---T	4320
Db	7775	GMSLTFTSVIRGTPPFVKVFKPSRELVPGESCNISLEDVFTLELFEVQPLSGDYSCIL	7834
Qy	4321	VSPHVGTCSEAOQKVTKATTVVRGLNVEALEGEALFECQLSQPEVAHTWLLDDEPV	4380
Db	7835	VTNDAGASCTTHLFVKEPATFVKRLADFVSFTGSPIVLEATVTGTPPISVSWIKDEYLI	7894
Qy	4381	RTSENASWPFENGLRHLLLLKNLR--PODSRCRVTFLAGDMVTSFAFLTVRGMLE---IL	4435
Db	7895	SQSERCSITWTERKSTILEILESTIEDYAOYSCLUENEGADICEALVSV---LEPPYFI	7950
Qy	4436	EPLKNAAVRAGAQAARFTCTISEAVPVGEASWYINGAAVQPDSDTWTVDGSHQALLRS	4495
Db	7951	EPELHEAVTCEPATLOCKY-DGTPETIRISWYKHTKRLSAPA-YKQPKFNWVASIVINK	8008

Qy	4496	AQPHHAGVTFPACR	-----DAVASARLTVL-----GLPDP-----	PEDAEVVA	4533
Db	8009	VDHSDVGE--YSCADNSVGAASSALVITIKERLPPFFARKLKDVIHETLGLFPVAPECRI	8066		
Qy	4534	HSSHVTLSWAAWMSDCGGGLCGVYREVKEGATGOWRLCHELVPGBCVVDGLAPGETYR	4593		
Db	8067	NGSEPLQVSW--YKOG-----VLLKODANLQTSFVHNANTLQLQTDQSHIGQ----	8112		
Qy	4594	FRVAAGFVGAGEPVHLPQTVRLAEPPKPYPP--QPSAPESRQVAAGEDVSLELVVAEAG	4652		
Db	8113	YNCSASNPLGTAS-----SSAKLILSHEVPPFPFDLKPVSVDLALGESGTFKCHVTGTAP	8167		
Qy	4653	EV1WHKMERIOPGRFEFVVSOGROMLVIKGFATDQGEYHC-----GLAQGSICPAANAT	4708		
Db	8168	IKITWAKDNREIRPGGNYKMTLVENTATLTVLKVKGAGQAGQYTCYASNIAGKDCSA----	8224		
Qy	4709	FQVALSPASVDEAPQPSLPPENAAQEGDLHLLWEALARKRMSREPTLDSISELPEBDGRS	4768		
Db	8225	-----HLGVQEPFR-----FIKLEPSRIVKQDEFTRYECKIGGS	8259		
Qy	4769	QRLPOEAEVAPDLSE--GYSTADELARTGDADLSHTSSDDSESAGTGPLSVLYLTKAGRPG	4827		
Db	8260	-----PEIKVLWYKDETEIQESSKFRMSFVDS-----VAVLE-----	8291		
Qy	4828	TSPLASKVGAPAPSVKPOQOQBPAAVRPPLDGLSTYKOLDGDSMD--KAANVKIQAAFKG	4885		
Db	8292	-----MENLSVEDSGDYTCBAHNAAGSASS-----	8317		
Qy	4886	YKYRKEMKQOQEGPMFSHTFGDTAQVGDALRLCEVVASKADVPARWLKOGVELTDCGRHHH	4945		
Db	8318	-----TSUKVKEPPIFRKKPHPIETLKGADVHLECELGQTPPFHVSYWKQRELSRGSKKYK	8373		
Qy	4946	IDQLGDCGTSLLTAGLDRADAGCVTCQVSNKFGQVTHSACVYVVGSES--EAESSSGGELD	5004		
Db	8374	I-MSENFLTSHILNVDAADIGHYQCKATNDVGSDT---CV---GSIALKAPPRFVKULS	8426		
Qy	5005	DAFRRAARRLHRLFRTKSPAESDEELFSADEGPAPBEEPADQWYREDEHFCIRFEA	5064		
Db	8427	DISTVVGKEV--QLQTTIEGAEPISVVMF--KDKGEIVRESDNWISY--SENATLQPSR	8481		
Qy	5065	LTEARQAV-----TRFOEMATLIGIVEIKLVQGGPRRVEMCISKETPAPVVPPEPL	5116		
Db	8482	VEPANAGKYTCQIKNDAGMOECFATLSVLEPATIVEK-----PESI	8522		
Qy	5117	PSLLTSDAAPVFLTELQNSQVQDGYPSVPCVVTGQPMPSVRMFKDKLLEEDHMYIN--	5175		
Db	8523	-----KVTTGDTCTLECTVAGTPELSTKWFKQKGLTSDNKYKISF	8563		
Qy	5176	EDQOQGHOLIITAVPADMGVYRCLAEWSNGVSTKAEURLVDLTSTDYDTAADATESSY	5235		
Db	8564	FNKVSQ--LKIINVAPSDSGVSYFQVNPVGVKDCSTASLQV-----	8602		
Qy	5236	PSAQYLSRSREQGETSTDEGQLPQVVEBLRDLQVAPGTRLAKFQKVGYPAPRLYWF	5295		
Db	8603	-----SDRTVPPSPFTRKLKETNGLSGSSVY--MECKVYGSPPISVSMF	8643		
Qy	5296	KDQPLTASAHIRMTGKKILHLEIISVTREDSGQVAAVYSNAGWGAAYSARLLVRGPD	5355		
Db	8644	HEGNEISSGRKYQTTLTDNTCALTVNMBESDGDYTCCTATNAGSDECSAPUTVR----	8699		
Qy	5356	PEEKPADSVHEQLVPPRMLERFTPKVKYKGSSTIFSQVKEGVPVTVHMLREAEARGVLW	5415		
Db	8700	-----EPPSFVQKPDMPDMVLTGTVNTFTSIVKGTFFPSVSMFKGSSEL----	8742		
Qy	5416	IGPDTPG--YTVASSAQOHSVLVLDVGRQHQGYTTCIASNAQOALCSASLHVGLPKVE	5473		
Db	8743	-----VPGDRCNVSLSDSVASLELFDVDTDSQSEYTCIVSNEAGKASCTTHLYTIKAPAFV	8798		
Qy	5474	EQ-----EKVKEALISTFLQGT---TQALISA-----QGLETASFADLGGORKEEPLAAKE	5520		
Db	8799	KRLNDYSIEKGKPLI-----LEGFTTGPPISVTVWKKNGINVT-----SORCNITTEKS	8849		
Qy	5521	ALGHLSLAEVGTBE-----FLOKLTSQITEMVSAKITQAK-----LOVPQGDSD	5564		

Db 8850 AI--LEIPSTVEDAGQVNCYIENASGK--DSCSAQILILEPPYFVKQLEPKVSGVDS 8905
Qy 5565 EDSKTPSPRGRS-----RPSSS-----IQES-----SSESE 5593
Db 8906 SIQCQLAGTPEIGVWYKGDTKLRPTTITVXMHFRNNVATLVFNQVDINDSGEYICAKENS 8965
Qy 5594 DGDARGEIDYIVVTADYPLPLGAEQDAITLREGQYVEV-----LDAAHPL--RMLVRTKP 5646
Db 8966 VGEVSASTF--LTVOEQKLPPSFSQRLRDVQETVGLPVVDFCAISGSEPISVSWYKDGKP 9023
Qy 5647 TKSSPSROGWSPAVLDRLRLKSPWGAABEPEFGEAVSEDEYKARLSSVIOELLSESEQ 5706
Db 9024 LKDSFN---VOTSFLDNTATLN-----IFKTRSLAQOYCSCTATNPIGASSAR 9070
Qy 5707 AFVEELQLOSHLQHLCRCHVPIAVAGQKAVI-----PRN 5743
Db 9071 LILTE-----GKNPPFDIRLAPVDAVAGESADFECHVTGTQPIKIVSWAKOSRE 9119
Qy 5744 VRDIGRHSFLQE-----LOQCDTDDVAMCFIKNQAAFOYLEFVLGRVQAESVVVST 5798
Db 9120 IRSQKYOISYLENSAHLTVLKVKGDS-----QYTCYAVNEVGKDSCTAQL 9167
Qy 5799 AIOEYKKYABEALLAGDSPPPPOHYLEQPVVERVORYOALLKELIRKARNONCA 5858
Db 9168 NIKE-----RLIPSPFTKRLSETVEET----- 9189
Qy 5859 LLEQAVVVSALPQRAENKLVHSLMENYPGTLALGEPIROGHFIVWEGAGAMP----- 5914
Db 9190 -----EGNSFKLEGRVAGSQPIVA 9209
Qy 5915 WKGNHRV-----FLFRNLHVICKRRDRSRTDTVSVFRNMKLSIDLNDQVEGDDR 5967
Db 9210 WYKNIEIQPTSNCEITEPKNNTLVLOVRKAGMDAGLYT-----CKVS-----ND 9254
Qy 5968 AFEVWQEREDSVRKVILLOARTAILKSSWV--KEICIGIQORLALPWRPDPFEEELADCTAE 6026
Db 9255 -----AGSALCTSSIVIKE-----PKKPPVFDQHLTPVTVS 9285
Qy 6027 LGETVKLACRVGTGPKPVISWYKDGKAVQVDPHHILIEDPDGSCALILDSLTGV----- 6080
Db 9286 EGEVQLSCHVOGSEPIRIQMLKAGREIK-----PSDRCSFSGASTAVLEIRDV 9335
Qy 6081 ---DSQWMCFAAAGNCSTLGKILVQVPPR-----FVNKVRASPFVEG 6122
Db 9336 AKADSGDYVCASNAGVSDTTKSKVTIKDKDAVAPATKKAADVGLFVSPQSIKRVVEK 9395
Qy 6123 EDAQFTCTIEGAPYQIRWYKDGALLTGNKFOTLSE-----PRSGLLVIVIRAASKE 6175
Db 9396 TTATPIAKVGGDPIPNVKW-----TKGKWRQINQGGRVFIHOKGDEAKLEIRDTTKT 9447
Qy 6176 DLGLYECELVNRLGSARASAEIRI-----QSPMLQ----- 6205
Db 9448 DSGLYRCVAFNEHEIESNVNLQVDERKKQEKIEGDLRAMLKKTPIILKKGAGEBEEIDIM 9507
Qy 6206 -----AQOCHRBQLVAAVEDTTLERADQ 6229
Db 9508 ELLKNVDPKEYEKYARMYGITDFRGLLOAFELLKQSQBEETHREIEBIERSERDEKEFE 9567
Qy 6230 EVTSVLAKLLGPKAPGSTGLTGPSCP-----RGAPALQ-- 6265
Db 9568 ELVSFIQORLSQTEPVTLIKDIENQTVLKONDAVFEIDIKINYPEIKLSWYKGTGKLEPS 9627
Qy 6266 -----ETGSQPPVTG-----TSEAPAVPPRPVQPLLHEGP 6295
Db 9628 DKFEISIDGRHTRLRVKNQCKQDQGNVRLVCGPHIASAKLTVIEPAMERHLQDVTLKEG- 9686
Qy 6296 EQEPEAARAEQWTP--IRMEGAAM-----PGAGTGELMDVHSHV---VRETTOR 6342
Db 9687 ---QTCMTQCFSVPNVKSE---WFRNGRILKPGGRHKTEVEHKVHKLTIADVRAEDQG 9739
Qy 6343 TYT--YQAIIDHT-----ARPPSMQVTEIDVQAGTGTAOFAEALIEGDPQSPVWYKDSV 6395

Db 9740 QYTCYKEDLETSABRLIEABPIQFTKRIQNIWVSEHQSATPECEVSFD--DAIVTWYKGPT 9798
Qy 6396 QLVDSITLSQOQEGTYSVLVRHVASKDAGVYTCLAO--NCGQVLCBAEL-----LV 6446
Db 9799 ELTESQYNFRNDGRCHYMTIHNVTDPDGBGVYSVIARLEPRGEARSTAEIYLTTKKEIKLE 9858
Qy 6447 IGGDNEPDS-----EKQSHRRKLHSFVEYKKEE 6473
Db 9859 LKPPDIDSDRVPITPMPIRAVPPEEIPPVVAPPIPLLLPTPEEKKPPPKRI-----EYTK- 9913
Qy 6474 IGRGVFVKRVQHKGNKILCAAKFIPLRSTRQAYRERDILALSHPLVTGLDQFET 6533
Db 9914 -----KAVKDAKAVVAKPKEMTPREBEIVKKVPPPTTLLIPAKAPEII-----DVSS 9959
Qy 6534 RKTLLILELCSSEEL---LDRLYRKGVTVEAEVKVVIQOLVEGLVHLHSHGVHLHLIKP 6590
Db 9960 KAEVKIMTITRKEVOKEAEVYKQAVHKEKRVFIESFEEDY-----ELEVPEP 10011
Qy 6591 SNILMVHP-----AREDIKI-----CDFGAQNIPTAELOFQOYGSPE--- 6628
Db 10012 YTEPFEOYVEEDDYEEIKVEAKKEVHEWEEDFEGQYVYEREBGYDE--GEEWEEA 10070
Qy 6629 FVSPETIQ--QNPVSEAS-----DIWAMGVISYLSLTCSPPAGSDR 6669
Db 10071 YQEREVIQVQKEVYESSHERKVPKAVPEKKAPPPKVIKPKVIEKIEKT--SRREEEKVQ 10129
Qy 6670 ATLLNVLEGRVSNSSMAAHLSEDAKDFIKATLQAPQAPPSAAQCLSHWFLKSPAE 6729
Db 10130 VTKVPEVSKKIVQKPSRTPVQBEVIEV-----KVPVHTKKMVISEEXKMFASHTEEE 10183
Qy 6730 AHFINTKQLFLRLARSQRSLMSYSILVMRSIPELLRGPDPSPSILGVARHLCRDTGGS 6789
Db 10184 VS--VTVEP-----QKEIVTEKIHV--AVSKRVEPPKVPPEL----- 10218
Qy 6790 SSSSSSDNLELAPAKAKSL--PPSPVTHSLPLHPRGLRPSASLSPHEAEASESTERAP-A 6847
Db 10219 --PEKPAPEEAVPVPKKEVPAPVPEV-----PKKVPPEKKVPVPKKEPAA 10267
Qy 6848 PPASPE-----GAGPPAAQOCVPRHSVI--RSLFYHQAGSPHEGA- 6886
Db 10268 PKKVPPEVPKVPPEEKIPIVPVAKKKEAPKAPKVPDEVKRVVTEEKIIVTQRESPPPAVP 10327
Qy 6887 -----LAPGSRHRHARRHLLKGGYIAGALPGLREPLMEHRV-----LE 6925
Db 10328 EIPKKVKVPBERKVPKKEEVEVPPPKVPALP--KCPVPEKAVVPVPVAKKAPPAEVS 10385
Qy 6926 EEAAREEQATLLAKASFETALRLPASGTHLAPCH--SHSLEHDS- 6970
Db 10386 KKTVBEKRPVABEKLISFAVQVRVTRHEVSAEBSWSYSEEEGVISIVYREEBREEE 10445
Qy 6971 -----TPRPSSEACGE---AORLPSAPSGGAPIRDMGHGPOGSK-----QLPSTG 7011
Db 10446 EAEVTEYVMEERPEEYVVEKLIHISKRVEAEPA-----EVTEREKKIIVLPKPIPAKI 10499
Qy 7012 GHFGTAQ--PERSPDPSWGO--PAPFCHPKQGSAPQBGCSPPHAPVAPCPGSPFPGSCK-E 7068
Db 10500 EEPKPAKVPKAPKVIPEKKVPAPV--PKKEKVPVPKVPPEEPK--KVPKPKVPVPKVIKME 10556
Qy 7069 APL-----VSSSPFLGOQOAPPAPAKASPDLQKMG--PGDISLPGRPKPGPCSS 7116
Db 10557 EPLPAKVTERRHQITQEBKVLVAVTKCEAPPKARVPEEPKRAVPEBKVLKLPKREE--EP 10615
Qy 7117 PGSASQASSQSSLLVAGSSQVGTGEPGLSDABGWTQE--AEDLSSTPTLQRPQOATM 7174
Db 10616 PAKVTEFRKRVKKEKVSIEAPKREPOPIKEVTIMEEKERAYTLEEAVSVQREYEEY 10675
Qy 7175 RKPFLGGRGGYAGVAGYGTTFAGDAGGMLGOGPMWARIAWVSOSEEBEQEBARAESQS 7234
Db 10676 EEDYKEFEPEYPTTEYDQY-----EYEBEREYERYEHE 10710
Qy 7235 EEQOEAREASPLQVARSARVPEVGRAP-----TRSSPE---PTPWEDIQVSLVQIRDLS 7286
Db 10711 EYITEPEKIPKVPPEEPVPTKPAKPAKVLKAVPEEKVVPV----- 10754

```
Qy 7287 GDAEADTISLDISEVDPAYLNLSLDYIKYLPFEFMIFRKVPKSAQPEPSPMAEBELA 7346
Db 10755 -----IPKLLKPPPK-VPEPPKK 10772
Qy 7347 EFPPTWPGELGHAGLEITE-ESEDV-----DALLAEAAVGRKRKWSGPSR 7394
Db 10773 VFEB-----KRISITKEKEQVTPAAKVPKRVVAEEKVPVPRKEVAP-- 10819
Qy 7395 SLFHPGR--HLPID-BPAELGLRERVKASVEHISRII-----7429
Db 10820 -----PVRVPEVPKELEPEEVAPEEV---VTHVEYLVEEEVYIHEEEFFITEEVVP 10871
Qy 7430 -----KGRPEGLEKEGPPRPKP 7446
Db 10872 VIPKVPVRPKVPPEEKVPVPVPPKKKEAPPKVP 10906

RESULT 11
Q8NH4 PRELIMINARY; PRT; 658 AA.
AC Q8NH4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Obscurin (Fragment).
GN Name=OBSCN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gautel M.S.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314904; CAC85751.1; -.
DR InterPro; IPR007110; IG-like.
DR Pfam; PF00047; 19; 1.
DR SMART; SM00408; IGC2; 3.
DR PROSITE; PS00835; IG LIKE; 4.
FT NON TER 1
FT NON TER 658
SQ SEQUENCE 658 AA; 72830 MW; CB37B453FCD4AE4A CRC64;

Query Match 8.2%; Score 3389; DB 2; Length 658;
Best Local Similarity 100.0%; Pred. No. 1.3e-101;
Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2075 RQVLRLVGLQAVEAREOGTATMEVOLSHADVDSWTRDGLRFQOGPTCHLAVRGMHTLT 2134
Db 1 RQVLRLVGLQAVEAREOGTATMEVOLSHADVDSWTRDGLRFQOGPTCHLAVRGMHTLT 60
Qy 2135 LSGLRPDSGLMVFAEGVHTSARLVLTVELPVFSRPLQDVVTTKEKVTLECELSRPNV 2194
Db 61 LSGLRPDSGLMVFAEGVHTSARLVLTVELPVFSRPLQDVVTTKEKVTLECELSRPNV 120
Qy 2195 DVRLKDGVELRAGKTWAIQAQACRSITIVRCFADQGVVYCDADHAQSSASVKVQGR 2254
Db 121 DVRLKDGVELRAGKTWAIQAQACRSITIVRCFADQGVVYCDADHAQSSASVKVQGR 180
Qy 2255 YTLIYRVLAEADAGEIQFVAENAESRAQLRVKELPVTILVRDLRKIAHEKRGVLECOVS 2314
Db 181 YTLIYRVLAEADAGEIQFVAENAESRAQLRVKELPVTILVRDLRKIAHEKRGVLECOVS 240
Qy 2315 RASQVRFWFGSQELQPGPKVELSDGLYRKLIIISDVHAEDEDTVTCAGDVKTSAQFFV 2374
Db 241 RASQVRFWFGSQELQPGPKVELSDGLYRKLIIISDVHAEDEDTVTCAGDVKTSAQFFV 300
```

```
Qy 2375 BEQSITIVRGLQDVYVMEPAPAFECETSPSVRPPKWLICKTVLQAGNVGLQEQTIVH 2434
Db 301 BEQSITIVRGLQDVYVMEPAPAFECETSPSVRPPKWLICKTVLQAGNVGLQEQTIVH 360
Qy 2435 RLMRLRRTCSMTGTPVHTVGKSRSSARLVSDIPVLTPLRPLEPKTGRQLQSVLSCDPRP 2494
Db 361 RLMRLRRTCSMTGTPVHTVGKSRSSARLVSDIPVLTPLRPLEPKTGRQLQSVLSCDPRP 420
Qy 2495 APKAVQVYKDDTPLSPSEKFKMSLEGQMAELRLRLMPADAGVYVRCQAGSAHSSTEVTV 2554
Db 421 APKAVQVYKDDTPLSPSEKFKMSLEGQMAELRLRLMPADAGVYVRCQAGSAHSSTEVTV 480
Qy 2555 AREVTVTGLODABATEEGWASPCSELSEDEEVNSLNGMPLYNDSFHEISHKGRHRTL 2614
Db 481 AREVTVTGLODABATEEGWASPCSELSEDEEVNSLNGMPLYNDSFHEISHKGRHRTL 540
Qy 2615 VLKSIQADAGIVRASSLKVSTARSLEVRVKPVVFLKALDDLSAEERGTALAOCEVSDPE 2674
Db 541 VLKSIQADAGIVRASSLKVSTARSLEVRVKPVVFLKALDDLSAEERGTALAOCEVSDPE 600
Qy 2675 AHVVVRKDGVLGSPDKYDFLHTAGTGLVVDVSPEDAGLYTCHVGSEETRARVRVH 2732
Db 601 AHVVVRKDGVLGSPDKYDFLHTAGTGLVVDVSPEDAGLYTCHVGSEETRARVRVH 658

RESULT 12
Q10465 PRELIMINARY; PRT; 7962 AA.
AC Q10465;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Elastic titin (Fragment).
GN Name=titin;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Berkeley; TISSUE=Skeletal muscle;
MEDLINE=96026330; PubMed=7569978;
RA Labelit S.; Kolmerer B.;
RT "Titins, giant proteins in charge of muscle ultrastructure and elasticity."
RL Science 270:293-296(1995).
DR EMBL; X90569; CAA62189.1; -.
DR PIR; I38346; I38346.
DR HSP; P11362; IEVT.
DR GO; GO:0030017; C:sarcomere; TAS.
DR GO; GO:0008307; F:structural constituent of muscle; TAS.
DR GO; GO:0006941; P:striated muscle contraction; TAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF02818; PPAK; 53.
DR SMART; SM00408; IGC2; 43.
DR PROSITE; PS00835; IG LIKE; 58.
FT NON TER 1
FT NON TER 7962
SQ SEQUENCE 7962 AA; 883018 MW; B85240533CBADE58 CRC64;

Query Match 8.0%; Score 3313; DB 2; Length 7962;
Best Local Similarity 21.5%; Pred. No. 7.8e-98;
Matches 1710; Conservative 1105; Mismatches 3287; Indels 1868; Gaps 306;

Qy 10 PRFTRPKAFVSVGKDATLSCQIVGNPTQVSWKEQQQPVTAGARFRLAQDGLYRLTI 69
Db 10 PAITPLQDVTSEGPQPCRCVSGTDL-KVSWYSKDKIKIPSRFFMTQFDYQLLEI 68
Qy 70 LDLALDGSQVVCARNAIGEAFPAVGLQVDAEACAEQ-----APHFLRLPT 117
Db 69 AEAPDEGTFTFVANNVAVGVSTANLSLEAPESILHERIEQEIEMKAAPVIKRKIE 128
```


Db	1995	PSFLVKGROQAIPDSTVEFKAILK--GTPPPKIKWFKDDVELSGPKFCFLEGSTSF	2051
Qy	2222	LTIYRCBFADQGVVCDHAHQASASVKVQGRVTTLIYRRVLAEDAGEIQFVAENAESRA	2281
Db	2052	LNLYSVDASTKGQYTC-----VINDVG-----SOSCTT	2080
Qy	2282	QLRWKELPVTVLRLPLRKIAEMKHRGVLECOVSRASQAQR--WPKGSQELQPGPKYELVS	2339
Db	2081	MLLVTEPPKFKVLEASKIVKAGDSRSLECKIA-GSPEIRVWVFRNEHELPAADKYRMT-	2138
Qy	2340	DGLYRKLIISDVHAEDEDTYCDAGDVKTSQAQFVEQSITIVRGLQDVVMPEPAWFE	2399
Db	2139	-----FIDSVAVIQMNKLTSGD-----FI	2160
Qy	2400	CETSI PSVRPPKLLGKTVLQAGGNVLEQEGTVHRLMLRRTCTMTGPHVHTVGKSRSS	2459
Db	2161	CEAQNPA-----GST-----SGT-----	2174
Qy	2460	ARLVVSDIPVVLTRPLEPKTGRELQSVYLSCDFRPAPK-AVQWYKODTPLSPSEKFKMSL	2518
Db	2175	-KVIVKEPPVFSSPPIVETLKNAB-VSLECELSGTPPEPVWVWYKDKRLRSKKYKIAS	2232
Qy	2519	EGQWAEURLIIRLPADAGVRCQA-GSAHSTEV-TVEARE-----VTVTGPLQ	2565
Db	2233	KNFHTSIHILNVDTSDIGYHCKAQNBEVGSQDCTCYTKLEPPRFVSKLNSLTWAGEPA	2292
Qy	2566	DAEATEBG-----WASFCELSHEDEEVEWSLNGMPLYNDSPHEISHKGRHRTLVLKS	2618
Db	2293	ELQASIEGAQPIFVQWLKEKEEVIRESNT-----RITFVENVATIQFAK	2337
Qy	2619	IQRADAG-----IVRASSLKYSTSARLEVRVKPVFLKALDLSABERGTLALQCEYS-D	2672
Db	2338	AEPANAKKICQIKNDGGM-R-ENMATLWV-LEPAVIVEKAGPMVTIVGETCTLECKVAGT	2395
Qy	2673	PEAHVWRKGVQGPSKDYDFLHTAGTGLVHVDVSPEDAGLYTC-----HVGSBETRAR	2728
Db	2396	PELSVEVYKDGKLLTSSQKHFPYNNKISSURLILSVEROQDAGTYTFQOVNNGVSKSTAV	2455
Qy	2729	VRVHDLHV--GITKRLKTMELVLEGSCSPCVLSHESASDPAMVTVGKTVGSSSRFOAT	2786
Db	2456	VDUSDRAPSPFTRLKNTGGVLGASCLICKVAGSPISVA-WFHEKTKIVSGAKYQTT	2514
Qy	2787	RQGRKYTLVREAPS DAGE---VVFSVRGLTSKASLIVRERPAALIKPLEDQWAPGED	2843
Db	2515	FSDNVCTQLNSLSDSDMGNYTCVAANVAGSDECAVLTVQEBPSPFVKBEPELVLPGKN	2574
Qy	2844	VELRCELSRAGTP---VHMLKDKRAIRKSOKYDVVCEGTWAMLVIRGASLKDAGEYTCV	2900
Db	2575	VTFTSVI--RGTPPFKNWFRGARELVKGDRCNIYFEDTVAEELFNIDISQGEYTCVV	2632
Qy	2901	EASKSTAS---LHVEBKANCFTEELTNLOVEE-KGTAVTCTKTEHPAATVTRKGLLEL	2955
Db	2633	SNNAQGSCTTRUFVKPEPA-AFLKRUSDHVEPEKSIILBSTYTGTLPTISVTKWKGDFNI	2691
Qy	2956	RASGKHOPSQEGLTRLTISALEKADSDTYTCIDIGQAQSAQLLVQGRRVHIIE-----	3009
Db	2692	TTSEKCNIVTETKTCILEILNSTKRDAGQYSCFI---ENAGRDVCGALVSTLEPPFYVT	2748
Qy	3010	DLEBDVVOEGSSATFRICRISPANYEPVHWFLDKTPPLHANE-----LNEIDAQPGGVHVL	3063
Db	2749	ELEPLEAAVGDVSLQCVQVAGTPEITVSWYKGDTKLRPTPEYRFTYFNNV-----ATL	2801
Qy	3064	TLQLALKLDSGTIYFEA---GQORASAAALRVTEK--PSVFSBELTDATITEGEDTLVVC	3117
Db	2802	VFNKVNINDSGEYTCKAENSIGTASSKTVFPRIQERQLPPSFARQLKDIBQTVGLPVLTC	2861
Qy	3118	E-TSTCDIPMCWTXKDGKTLRGARSARCOLSHGHEHRAQLLITCATIQDQGRYKCEAG---	3173
Db	2862	RLNGSAPIQVWYRDGVLLRDHENLQTSFVDNVATLKILOTDLSHSGOYSCSASNPLGTA	2921
Qy	3174	SSSIVRVHARPVR-----FQBALKDLEYLEGGAATLRCVLSSVAAPVK--WCYGNVLRPG	3227


```
QY 5781 YLEFLVGRVQAEVVVSTAIQBYFYKYAEALLAGDPSPPPPLQHYLEQVERVQYQ 5840
Db 181 YLEFLVGRVQAEVVVSTAIQBYFYKYAEALLAGDPSPPPPLQHYLEQVERVQYQ 240
QY 5841 ALLKELIRNKARNRQNCALLEGAYAVSALPQAEKLVHVSIMENYPTGLGALGSPRQ 5900
Db 241 ALLKELIRNKARNRQNCALLEGAYAVSALPQAEKLVHVSIMENYPTGLGALGSPRQ 300
QY 5901 HFIVMEGAPGARMKGNHNVFLFRNLHVICKPRDSRTDTVSFVFRMMLSSIDLND 5960
Db 301 HFIVMEGAPGARMKGNHNVFLFRNLHVICKPRDSRTDTVSFVFRMMLSSIDLND 360
QY 5961 QVEGDDRAFEVWQEREDSVRKVLLQARTAIKSSWVKEICGQORLALPVWRPDPFEEL 6020
Db 361 QVEGDDRAFEVWQEREDSVRKVLLQARTAIKSSWVKEICGQORLALPVWRPDPFEEL 420
QY 6021 ADCTAELGETVKLACRVGTGPKPVIISWYKQKAVQVDPHHILIEDPDGSCALILDSLTGV 6080
Db 421 ADCTAELGETVKLACRVGTGPKPVIISWYKQKAVQVDPHHILIEDPDGSCALILDSLTGV 480
QY 6081 DSGQVMCFAAAGNCSTLGLKLVQVPRFVNKVRASPFVGEDAQFTCTTIEGAPYPOIR 6140
Db 481 DSGQVMCFAAAGNCSTLGLKLVQVPRFVNKVRASPFVGEDAQFTCTTIEGAPYPOIR 540
QY 6141 WYKDGALLTTGNKFQTLSEPRSGLLVLVIRAAKEDLGLYCELVNRLGSRASAEALRIQ 6200
Db 541 WYKDGALLTTGNKFQTLSEPRSGLLVLVIRAAKEDLGLYCELVNRLGSRASAEALRIQ 600
QY 6201 SPMLQAQOCHREQLVAADVDTTLR-----ADQEVTSVLKLLGPKAPGSPSTGDLT 6252
Db 601 SPMLQAQOCHREQLVAADVTEQETKVPKKTIVIEETITTVKVSQRQR-SPSKSPSR 659
QY 6253 GPGPCPRG-----APAL-----QETGSP--PVTGTSAPAVPRV---- 6286
Db 660 SPFRCSASPLRPLGLLAPDLLVLPNGQPRRPEAPGQPVVPTLYVTEAEHSAHPALPGLS 719
QY 6287 -PQPLLHEGPE-----QBPEAIARAQEWTVFIRMEGAAW--PGAGTG 6325
Db 720 GPQPAWVEVEETIEVRVKMKMGPGQVSPTE--VPRSSSGHLFTLPGATPG 767

RESULT 14
Q72120
ID Q72120 PRELIMINARY; PRT; 8081 AA.
AC Q72120;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Uncoordinated protein 89, isoform b.
GN Name=unc-89; ORFNames=C09D1.1;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.,";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
PA Du Z., Le T.T., Wilson R.;
RT "The sequence of C. elegans cosmid C09D1.1";
RN Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
```

```
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RN Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Contains 1 PH domain.
DR EMBL; AF003131; AAP68958.1; -.
DR HSSP; O01761; 1FHO.
DR WormBase; WEGene00006820; C09D1.1.
DR WormPep; C09D1.1b; CE34251.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000577; FGGV_kin.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; Ig_c2.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR011009; Kinase_like.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR007850; RCDSD.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR011511; SH3_2.
DR Pfam; PF00041; fn3_2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00069; Pkinase; 2.
DR Pfam; PF05177; RCDSD; 5.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF07653; SH3_2; 1.
DR ProDom; PD000001; Prot_kinase; 2.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGC2; 24.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00010; DH_2; 1.
DR PROSITE; PS00933; FGGV_KINASES_1; UNKNOWN_1.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS00835; IG_LIKE; 50.
DR PROSITE; PS00003; PH_DOMAIN; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00002; SH3_1.
SQ SEQUENCE 8081 AA; 894245 MW; 67C804953CF62228 CRC64;

Query Match 7.7%; Score 3182.5; DB 2; Length 8081;
Best Local Similarity 21.1%; Pred. No. 1.3e-93;
Matches 1891; Conservative 1190; Mismatches 3350; Indels 2549; Gaps 375;

QY 27 ATLSGCVGNPTPOVSWKDDQOPVTAGARFLAQDGD-LYRLTILDALGDSQYVCRAR 85
Db 565 AKMCLKVTGYPLDITWYKDDVQLHDEDRHTFYSDEDDGFAMTIDPVQVTDTRGYTGMAT 624
QY 86 NAIGEAFAAV--GLQVDAEACAEQAPHLRPTTSIRVREGSEATFRCRVGSGPRPAPS 142
Db 625 NEYGQASTSAFFRVLKVEKEA---PPAFVTKLRDKEGVDIDFECEVEGWPELV 680
QY 143 WSKDGRRLGEPDGPVRVVEELGEASALRTRARPRDGGTYEVRANPLGAASAAALVVD 202
Db 681 WLVDQDLRPSHDFRLQYD--GQTAKLEIRDAQDDTGVYTVKIQNEFGSIESKAELEFVQ 738
QY 203 SDAADTASRPGTSTAALLAHLQRRREARAEAPASPPSTGTCTCTVTEGKHARLSCYVT 262
Db 739 ADPDKNHVAP-----EFQATIEVECD-----EGEVRFKSVIT 772
```


Db 2504 ITSDGHYKIVEEDGSLKLSIQTKLEDKGRITTIKAESEFGVAECASLGV-----VK 2556
QY 2260 RRVLAEDAGEIOFVAENAESRAQLRVKGLPVTLRPLRDKIAMEKXRGVLECOV---SRA 2316
Db 2557 GRPMAKPA-----FQSDIA-----PINLT-----EGDTLCKLLITGDP 2590
QY 2317 SAQVWFVKSQBLQPGPKYELV-SDGLYRKLIISDVHAEDEDTYTC-----DAGDVKTSAQ 2371
Db 2591 TPFVKWYIGTQVUCATEDEISNANGVY-TMKIHGVATDMTGKIKCVAYNKAGEVSTBG- 2648
QY 2372 FVVEQSITIVRGLQDVVMEPAPWFE---CETSI-----PSVRPPK 2411
Db 2649 -----PLKVAPIPVEFETSLCDATCREGDTLKURAVLLGPEPEPVV---S 2690
QY 2412 WLLGKTVLQAGNVGLEOEGTVHRLMLRRTCTMTGPHVFTV---GKSRSARLWVSDI 2467
Db 2691 WYVNGKLEESQNIKIHKSEGTYYTVIKDIICDYSQVVCEAINEYGKATSEATL----- 2745
QY 2468 PVVLTTRPLEP-----KTG-RELQSVVLSCDFRPAPKAVQWKDDTPLSPSEKFK 2515
Db 2746 -LVLPGRGPPDFLEWLSNVRAARTGTVKHVKVFTGDPKP---SLTWYINNKEILNSDLXT 2801
QY 2516 MSLEQOMAEALRILRLMP-----ADAGVVRCOA-----GSAHSSTEVTVREARE 2557
Db 2802 IVTDDKTSTLTINSFNPVDVHVEBIEICKAENDAGEVSCFANNITVTSOMFSESESEAQABE 2861
QY 2558 -----VTVTGP-----LQDAEATEEGWASFCELSHEDBE---VE 2589
Db 2862 FVGDDLTDESIREMHRTPTPVMAKFIITKIOTKAKKGHSAPFECVV---PDTYGVCKK 2919
QY 2590 WSLNGMPLYNDSFHEISHKGRHT-----LVLKSTORADAG-----IVRASSLKV 2634
Db 2920 WLKDGKEI-----ELIARIVQTRTGPGEHITQELVLDNVTPEDAGKYTCIVENTAGKD 2973
QY 2635 STSARLEV-----RVKPVFLKALDLSABERGTLALOCEV-SPEAHVVRKDGQV 2695
Db 2974 TCEATLVIESLEKKEKKAPEFIVALQDKTKTSEKVLVECKVIGBPKVSWLHDKNT 3033
QY 2686 LQPSDKYDFLRTAGTGLVHVDSPEDAGLYTCHV---GSEETRARVRVHDLHVGITXR 2741
Db 3034 I-TQESITVESVEGVERVITTSBELSHQGYCTIAENTEGTSKTEAFITVQGEAPVFTKE 3092
QY 2742 LKTHEVLEGCSCFECVLSHESADPAWTVGGKT-----VGSSSRF-----QATROGRK 2791
Db 3093 LONKELSTIGEKLVLSCVSGSPQPHVDYFYSFETTKVETKITSSRIAIEHDQNTHWRM 3152
QY 2792 YI-LVWREAPSDAGEVVVFSVRGLTSKASLIVRERPAALIKPLEDQWAPGEDVELRCEL 2850
Db 3153 VISQITKEDIVSYKAIATNSIGTATSTSKITTKVEAPVFEQGLKKTSGVKEKEIKMEVKV 3212
QY 2851 SRAGTPVHVKDKRAIRKSQKYDVV---CEGTMAMLVIRGASLKDAGBYTCVEASKSTAS 2908
Db 3213 GGSAPDVEWFKDDKPVSEDGNHMKKNPETGVFTLVVKQAATTDAGKYTAKASNPAGTAE 3272
QY 2909 LHVEKAN-----CFTEEL--TNLQVEEKGTAFTCTKTEHPAATVTRWKGLELBRASGK 2960
Db 3273 SSAAEAVTQSLKPTFVRELVTTVKINETATLSVTYKGV-PDPSVEWLKDGQPVQVQDSS 3331
QY 2961 HQPSQ-BGL-TLRLTISALEKADSDTYTC-----DIGQAQRAQLLVQGRV--HIIBDLE 3012
Db 3332 HVIAKVEGSGYSITIKDARLEDSGKYACRATNPAGEAKTEANFVAVVKNLVPFEFVEKLS 3391
QY 3013 DVDVQEGSSATFRCRISPANYEPVHVFMDKTPHLANELNEIDAQPG-GYHVLTLRQALAK 3071
Db 3392 PLEVKEKESTLTSVKVVGTPPEPSVEWFKDDTPISIDNVHVIOQTAVGSPFSLTINDARQ 3451
QY 3072 DSGTIYF-----EAGDQASALRVTEK--PSVFSRELTATITEGEDLLTIVETSTCDI 3124
Db 3452 DVG-IYSCRARNEAGEALTANFGIIRDSIPPEFTQKRLPVEBREQETDLKVTVTGTPV 3510
QY 3125 P-MCWTXDKGTLRGARSARCOLSH-----BCHRAQLLITGATLQDSGRYKC-----BAGGA 3172
Db 3511 PNVEWFKDDKPI-----NIDNSHIPAKDEGSGHHT-LTIKQAGEDGVYTCATNEAGEA 3565

QY 3173 CSSSIVRVHAR---PVRFOEALKOLEVLEGAATLRCVLSVAAP-VKWCGYNNVLRPGD 3258
Db 3566 KTTANMAVQSEIEIAPL-FVOGLKPYEVEQKPAELVVRVEGKPEPEVKNFKDGVPIALDN 3624
QY 3229 KYSLRQSG--AMLELVVRNLRPDQSGRYSC-----SFGDQTTSATL-----TVTAL 3272
Db 3625 QHVIEKKGNGSHTLVIKDTNNADFGKYTCOATNKAGKDETVGBELKIPKYSPEKQTAABE 3684
QY 3273 PAQFIGKLRNKEATEGATATLRCELSKTA--PVEWRKGSFETLRDGDYCYLR--QDGAMCE 3328
Db 3685 KPUIFIEPLKETFAVEGDTVVLECKVKNKESHPOIKFFKNDQPVFIEGQHMQLVELEDGNI-K 3743
QY 3329 LQIRGLAMVDAAEYSC-----VCGEERTSASLTI----- 3357
Db 3744 LTIQNAKEDVGAYRCEAVNVNAGKANTNADLKQPAKVVEHVTDESGQLEEBEQFTVVG 3803
QY 3358 -----RMPAHFPIGRHQESIEGATATLRCELSKAAP---VEMRK-GRESLRDG 3403
Db 3804 DTASSKTDTCRGAP-BFVELLRSCVTVEKQOAILKCKV-KGEPRPKIKMTKEGKEVEMSA 3861
QY 3404 DRHSLRQDGAVCCELQICGLAVADAGEYSCVCGEERTSA-----TLTVKALP-----AK 3451
Db 3862 RVRAEHKDDGTLTLTDFDNTQADAGEYRCEAENEYSAMTEGPIIVTLEGAPKIDGEAPD 3921
QY 3452 FTEGLRNEEAVEGATAMLWCELS-KVAP-VEMRKGPENLRDGDYILR--QEGTRCELOI 3507
Db 3922 FLOPVKPAVVTVGETAVLECKISGKPKSPVKYKNGEELKPSDRVKIENLDDGTQ-RLTV 3980
QY 3508 CGLAMADAGLYLVCQOE---RTSATLTIHALPAR-----FIEDVKNQEAAREGATAVLQC 3559
Db 3981 TNAKLDMDBYRCEASNEFGDVMSDVTLTVEK-PAQVAPGFFKELSAIQVKETETAKFEC 4039
QY 3560 ELNSAAP-VEMRKSETLRDGR--YSLRQDGTCKCELOIRGLAMADTGEV---SCVCGQ 3612
Db 4040 KVSGTDPDKVMFKDGTPLKEDKRVHESTDDGTQ-RLVIEDSKTDDQGNRIEVSNDAGV 4098
QY 3613 ERTSAMLTVRALP---IKFTEGLRNEEAETEGATAVLRCCEL-SKQAPVEMWKGHETLRDGD 3668
Db 4099 ANSKVPLTV-VFSETLKIKKGLTDVNVITQGTCKILLSVEVEGPKTVKMYKGTETVTSQ 4156
QY 3669 RHLRQ-DGARCBLQIRGLVAEDAGLYLWCQKE---RTSAMLTVR-----AMPSKFI 3717
Db 4157 TTKIVQVTESEYKLEIESAEMSDTGAYRVVLSTDSFSEVSATVTVTKAAKISLPS-FK 4215
QY 3718 EGLRNEEAETEGDTATLWCEL-SKAAPVEMRKGHETLRDGRHSLRQDGSRCCELQIRGLAV 3776
Db 4216 KGLADQSVKGTPLVLVEIEGPKDVVKYKNGDEIKDGKVEDL--GNGKYRLTIPDFQE 4273
QY 3777 VDAGEYSCV---CGQERTSATLTVRALPARFIEDVKNQEAAREGATAVLQCEL-SKAAPV 3831
Db 4274 KDVGESYVTAANEAGEIESKAKVNVSAKP-EIVSGLVPTTVKQGETATFNVKVPKGVKGV 4332
QY 3832 EWRKGSFETL-----RGDRYSLRQDGTGRCELOIHGLSVADTGEYSCV---CGQERTS 3880
Db 4333 KMYKNGKEIPDAKTDKONGDGSYS-----LEIPNAQVEDAADYKVVVSNDAAGDADS 4383
QY 3881 ATLTVRAP-----OPVPREPLQSLQAEBSGTATLQCELSBPTATVWVSKGLQLOANG 3933
Db 4384 AALTVKLADGDKVKVPEIVSGLIPTTVKQGETATFNVKVPKGVKQVKYKNGKEI-PNA 4442
QY 3934 RREPRIQGTAEALVLQDLOREDTGEYTCGSOA-----TSATLTVTAAVPRFLRELOHQE 3989
Db 4443 KAKDNGDG-SYSLEIPNAQLDDTADYKVVVSNDAAGDADSAAALTVKLPGIAIVKGLEDAE 4501
QY 3990 VDEGGTAHLCCELSRAGASVEMRKGSLOLFPKCAKQWMDG-AAEALLVRGVQEODAGDY 4048
Db 4502 VPKGKAVLQVETNKPKPKIKWYKNGKEITPDKAQPGSDGNKPOLPVPDAGDDDAEY 4561
QY 4049 ----TCDTGT--QSMASLSVRVP--RPKFKTLRQSLQEQTGDI-ARLCCQLSDAESGAVV 4100
Db 4562 KVVLTDEDGNTADSSCALTVKLPKPEKPI---IKGLEDOVVISGSPIKLEIETSGSPKT 4618

Db 6158 NKHFVNHMPTIFDCLVVGHPAPEVWFHNGKKIVPGGR1K-IQSCGGSHALIILDTTLE 6216
QY 6176 DLGLYECELNLRLGARSAAELRIQSPMLQAOEQCHREQLVAAVEDTTTLERADOEVTSVL 6235
Db 6217 DAGEYVATAKNSHGSSASSAVLDVTVPFL-----DSIKFNCEIDVTVPYL 6260
QY 6236 KRLLG-----PKAPGSTGDLTGG-----PCPRGAP----- 6262
Db 6261 TEEYGFKKLNTASLTPPDRGPFKEVTHGYLTLSWIPTRAPPRYPQVYVIEIRELPE 6320
QY 6263 ---ALQETGQPPVT-----GTSE-APAYPP-----RVPOPLIHE 6293
Db 6321 KOWSLLEYNIPEVCKVRNLELKSQYFVRAENYIGISDPSPASPPSRLMAPPOQVDFDR 6380
QY 6294 -----GPEQEPBAIARAQE-----W-----TVPIRMEG----- 6316
Db 6381 RTNKVPIPLDPAEAKALDMRYSEQACAPFSGVVEKRYCAENDTLIIVLNVSGFFPD 6440
QY 6317 AAWPAGAGTGLLWDVHSHVVRRTTORT-----YTYQAIIDHTVA----- 6354
Db 6441 IKWKFRG-----WDI-----DTSSPTSKCKVYTYGSGSETTLAITGFSKENVQYQCPAK 6489
QY 6355 -----RPSMOVTIEDVQAOIGTAQ-----FEAIIEGDPQPSVTHYKDSV 6395
Db 6490 NDYGDAQQNIMVDLATRPNFIQPLVN-----KTFSSAQPMRMDVRVDGFFPELKWMEWR 6545
QY 6396 QLVDSRLSQQOQGT-TYSLVLRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGDNBDP 6454
Db 6546 PIVESSRIKFVQDGPVLCSLIINDPMWRDSGIYSCVANADGAQTCTVTVEABGDYND 6605
QY 6455 SEKQSHR-----KLUHSFVEVKEETGRGVFGVKKRVQHKGNKILCAAKFIPILRSRTRAQA 6509
Db 6606 VELPRRRVTIESRRVRELYEISE-----KDEKL--AAEGAPFRVKEKATG 6648
QY 6510 -----YREDILAALSHPLVTGLDQFETRKTLILILELCSSEELLDRLYR 6555
Db 6649 REFLLAQLRPIDDALMRHVDIHNSLDPGIVQHRVLRDEKALVVFDD--NANSTIDGLSS 6706
QY 6556 -----KGVVTEAEVKYIQQVLVEGLHLYSHGVHLHDIKPSNLMVHVPAREDIK 6604
Db 6707 LAHPGVEIAEPKGVNRETQVRFVRLALLKALHMDLRIALHDLRPEITILL--QDDKLLK 6763
QY 6605 ICDFGFAQNIPTAELOFQOYSGSPFVSEIITQONPVSEASDIWAMGVISYLSLTCSSPPA 6664
Db 6764 LADFGQARLLRGLITGBIKSGSPFVSEIVRSYPLTLATDMWTGVLTYVLLTGLSPFH 6823
QY 6665 GESDRATLLNVLEGRVSWSSMAAHLSEDAKDFIKATLQAPQAPRPSAAQCLSHPW--- 6721
Db 6824 GDNDETLANVDSQCFD--SSPL--GNFSYDAGDFVKLLTEIPVSRLLVDEALDHPWINDE 6881
QY 6722 -LKSNPABEEAHFINTKQLKFLARSRWORSIMSYKSLVWRS-----IPELLRGP----- 6770
Db 6882 KLKTEP-----LSADTLREFKYQHKWLE-----RRVFVQQTPEQILEALIGPATQAQ 6930
QY 6771 PDSFSLGVARHLR-----DTGSSSSSSSSSSONE- 6799
Db 6931 QNAVAPGRPAEIIYDLRIQPKPPPTVEYVQPRKEHPFFIDFEGQLIDGDAFRPE 6990
QY 6800 ----LAPPARAKSLPPSP-----VTHS-----PILHPRGFL 6826
Db 6991 GTGFEGRHPQPPQIPQPPQRNQAADSRHREQQPHQGOQORIPVDQYGRPLVDPR-YL 7049
QY 6827 RPSASLPEEABASERSTAPAPSPAGAGPAPACGCVPRHSVIRSLFYHQAGSPSEHGA 6886
Db 7050 NDPSHRPSLLDAPFYVDKYGNPNVHFDFKYGRPMA-----PONLEKRRKLIPQDKGTPSHS- 7104
QY 6887 LAPGSRHRPARRRHLKGGYIAGALPG-----LREPLMEHRVLEEAAR-----E 6931
Db 7105 -KKEKTHQVATPIL-----ASPGDQOQOQKI PMRMINGERREREIEBIANRILSDIS 7155
QY 6932 EQATLLAKAPSFETALRLPASGTHLAPGHSLSLHSDSPSTPRSPSEACGAEQRLPSAPSG 6991
Db 7156 EEGSIAGSLASLED-FEIP-----KOFQVEASEPSTPTLTPE----- 7191

QY 6992 GAPIRDMGHPOGSKQLPSTGCHPGCTAQPERSPDPSWGPQAPFCHPKGSAPOEGCSPHP 7051
Db 7192 -VTIRE-----TIKPTPSPPTSPKSPVP----- 7214
QY 7052 AVAPCPPFPSPGSGKEAPLVSPSPFLGQOQAPPAPAKAS-----PPLDSKM--G 7099
Db 7215 -----QPQGLLI PAKVTYSDSILAGLPAADKKVLEDA 7246
QY 7100 PGDISLP-GRP--KPGPCSSPGSASQASSOVSLRVGSSQVGTPEPGPSLDAEGWTOEAE 7156
Db 7247 ENDPSSIIPVGAFLFLEGLHGSDDLTTDTTSASGL--IKVTSPLAINLSPNPK-----SPRS 7298
QY 7157 DLSDSTTFLORPOQATMRKFSGLGRCGYAGVAGYGTAFAGDAGGMLGOGPMWARIATA 7216
Db 7299 TPGTKSPVVLSPROHEHNEVLIIATKRG-----KPGFLPPGEL----- 7335
QY 7217 VSQSEEBEQEABEABESQSEEQEABESPLFQVSAREPVEVGRAPTRSSPEPTWEDIGQ 7276
Db 7336 ---AEDIDDEDAFMDDRKKQ----- 7352
QY 7277 VSLVQIRDLSGDAEAD-----TISL-DISEVDP--YMLSDL-----YDIKY 7317
Db 7353 ---VKPKDHGENDFKDEKERLEKDKNRRTVNLDDLKYPERSAFYKDDSDFGHPGYDIDA 7409
QY 7318 LPPE--FM1----FRKVPKSAQPEPPSPMAEELAEPEPTWMPWPGELG-PHAGLEITEE 7370
Db 7410 TPWDSHVQIIGPDTYLLMAARGAANRSRVNRYREELFGMGAPTVK-QGFLGVNRNDITVRER 7468
QY 7371 SEDVDALLABAAVGRKKNSSPSRSLFHPFGRHLPLDEPAELGLRERVKASVHSHIRLK 7430
Db 7469 RRYTD-ILRETTQGLEPKSHEQSTAL-----LQKAPSATAIERIKADIEKVTPC-- 7516
QY 7431 GREGLEKEGPPKKCLGLASFRLSGLKSWDRAPTFIRELSDETVVLVQSVTLACQVSAQP 7490
Db 7517 -----ATKKNDDGTF-----AFIFTARLDVLRKNQPAIFCAVSASP 7555
QY 7491 AAQATWSKOGAPLESSRVLISATLKNFOLLTLVLVVAEDLVVYTCVSNALGTVTVTGV 7550
Db 7556 APKVTWDFQKILLESNDRVITIEQD--NNVARLILNHAAPYDLGEVYCTAINEYGTDKSSCR 7614
QY 7551 LRKAERSSSPCPDIOGEVVAD-GVLLVWKPVEYGP-----VTYIVQCSL-----EGGSW 7599
Db 7615 LISGETPSRGRPE-AELSDSTEIFIQWEAPE--GPTYLEGITYRLEYRVAGPNHDGDPW 7671
QY 7600 TTLASDIFDCCYLTSLKSRGTYTFTACVSKAGMGYS-----SPSEQV--- 7644
Db 7672 ITVSEKIDDESIVVKHLSPLGIYQFRVTAQNGFGLGPLSSRSRIVQTHGKGAQKLQIDVL 7731
QY 7645 -----LLGGPSHLASEBESQGRSA-----QPLPSTKTFAFOTQ 7677
Db 7732 KSEIRLNVVMPQKSTNQLGGISE-ESEEDSEARTANEDMKSNLQLOTDDPTGRFOIGGL 7790
QY 7678 IQGRFSVVRQWKEKAS-GR-A-LAAKIIIPYHPKDKTAVLREYALKGLRHPHIAQLHAAY 7735
Db 7791 KFKGRFSVIRDAVDSTTEGHAHCAVKI--RHPSSE--AISEVESLRDGOHENVORLIAAF 7846
QY 7736 LSPRHVLVILELCSGPELLPCLAEARASYSESKVDLWQMLSATQVYLHNQHLHLDRSE 7795
Db 7847 NNSNFYLLSERLY-EDVFSRFVNDYTYEEQVALTMRQVTSALHFLHFKGTAHLVDNPH 7905
QY 7796 NMIITEYN--LLKVVLDLGNQAQSLQSEKVLPSDKFKDYLET--MAPEL-LEGQAVPQTDI 7850
Db 7906 NIMFQSKRWVVKLVDFGQAQKVS-GAVKPD-----FDTKWASPEFHIPETPVTVQSDM 7959
QY 7851 WAIGVTAFTMLAEYVPSSEGAARDLQRLKGLVRLSRVYAGI-----SGGAVAFURSTL 7905
Db 7960 WGMGVVTFCLAGFHPFTSE--YDREEEKENVINV-KCDPNLIPVNASQECLSFATWAL 8016
QY 7906 CAQPMGRPCASSCLOCPMLTTEEGPACSRPAPVTFPTARL 7945
Db 8017 KKSPPVRMETDEALSHKFLSSDPSMVRRESIKYSASRLR 8056

RESULT 15

Q8NHN8 PRELIMINARY; PRT; 646 AA.
AC Q8NHN8;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Obscurin (Fragment).
GN Name=OBSCN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RN [2]
RA SEQUENCE FROM N.A.
RA Gautel M.S.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ3114898; CAC85746.1; -.
DR HSP; Q10466; 1BPV.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 3.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00408; IGC2; 1.
DR PROSITE; PS50853; FN3; 1.
DR PROSITE; PS50835; IG_Like; 4.
FT NON_TER 1
FT TER 1
SQ SEQUENCE 646 AA; 70258 MW; B04F5B0084323D68 CRC64;

Query Match 7.5%; Score 3113; DB 2; Length 646;

Best Local Similarity 93.5%; Pred. No. 1e-92;
Matches 605; Conservative 16; Mismatches 24; Indels 2; Gaps 1;

Qy 421 NLLRKLPRKTAVRVGDGTAMFCVELAVPGPVHMLRNQEEVAVAGGRVAISAEGRHTLTIS 480
Db 1 NLLRKLPRKTAVRVGDGTAMFCVELAVPGPVHMLRNQEEVAVAGGRVAISAEGRHTLTIS 60
Qy 481 QCCLDEVGVAFMAGDCOTSTRFCVSAAPRKPLQPPDPVPVVKARMESSVILSWSPPHGE 540
Db 61 QCCLDEVGVAFMAGDCOTSTRFCVSAAPRKPLQPPDPVPVVKARMESSVILSWSPPHGE 120
Qy 541 RPVTIDGYLVEKKKLGTYTWIRCHEAEWVATPELTVDVABEGNFPQPRVSALNSFGQSPY 600
Db 121 RPVTIDGYLVEKKKLGTYTWIRCHEAEWVATPELTVDVABEGNFPQPRVSALNSFGQSPY 180
Qy 601 LEFGTGHAPLAVRTPLKAVQAVEGSEVTFSDLTIVASAGWFLDQALKASSVYEH 660
Db 181 LEFGTGHAPLAVRTPLKAVQAVEGSEVTFSDLTIVASAGWFLDQALKASSVYEH 240
Qy 661 CDTRHTLTIREVPASLHGAOLKPVANGIESSIRMEVRAAPGLTANKPPAAAAAREVLARL 720
Db 241 CDTRHTLTIREVPASLHGAOLKPVANGIESSIRMEVRAAPGLTANKPPAAAAAREVLARL 300
Qy 721 HEEAQLLAELSDQAAAVTWLKDGRTLSPGPKYEQVQASARRVLLVRDVARDDAGLYECVS 780
Db 301 HEEAQLLAELSDQAAAVTWLKDGRTLSPGPKYEQVQASARRVLLVRDVARDDAGLYECVS 360
Qy 781 RGGRIAYQLSVQGLARFLHKMAGSCVDVAVAGGPAQFECETSEAHVHVHWYKDGMELGHS 840
Db 361 RGGRIAYQLSVQGLARFLHKMAGSCVDVAVAGGPAQFECETSEAHVHVHWYKDGMELGHS 420
Qy 841 GERFLOEDVGTGRHRLVAATVTRQDEGTYSYCRVGEDSVDFRLRVSEPKVFAKEQLARRKL 900
Db 421 GERFLOEDVGTGRHRLVAATVTRQDEGTYSYCRVGEDSVDFRLRVSEPKVFAKEQACREV 480

Qy 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMCATGCTTRRLVVQAGQADAGEY 960
Db 481 QAEVGASATLSCEVAQAQDMETWYKDGKLSSSSKVHVEAVGCMRRLVVQVQVQADSGEY 540
Qy 961 SCEAGGQRLSPHLDVKPKPVFAKDQVAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
Db 541 SCEARGQVSPRLD--EPKVVFAKEQLARRKLQAEAGASATLSCEVAQAQTEVTWYKDGK 598
Qy 1021 KLSSSLKVHVEAKGCRRLRVVQQAQKTDAGDYSCEARGQVSPRLHI 1067
Db 599 KLSSSLKVCMCATGCTTRRLVVQAGQADAGEYSCAAGGQRLSPHLDV 645

Search completed: March 18, 2005, 17:58:56
Job time : 861 secs

THIS PAGE BLANK (uspto)

QY	181	GGCAGCTACAGTGTGAGGTGAGAGGCTTCCAAAGACACAGCAGCTCCAGTGTGAAGA	240
Db	988	GCCGAGTACAGTGTGAGGTGAGGCTTCCAAAGACACAGCAGCTCCAGTGTGAAGA	1047
QY	241	AAAGCAACTCTTACACAGAGAGAGCTGACCAATCTGCAGAGGTGAGAGAGAAAGCAGCT	300
Db	1048	AAAGCAACTCTTACACAGAGAGAGCTGACCAATCTGCAGAGGTGAGAGAGAAAGCAGCT	1107
QY	301	GTGTTACGTGCGAAGACGAGACCCCGCGCACAGTGAACCTGGCGCAAGGCGCTCTTG	360
Db	1108	GTGTTACGTGCGAAGACGAGACCCCGCGCACAGTGAACCTGGCGCAAGGCGCTCTTG	1167
QY	361	GAGCTACGGGCGCTCAGGGAGACACAGCCCGAGGAGAGGCGG	401
Db	1168	GAGCTACGGGCGCTCAGGGAGACACAGCCCGAGGAGAGGCGG	1208
RESULT 2			
AK097489			
LOCUS			
DEFINITION	AK097489	2554 bp mRNA linear	PRI 30-JAN-2004
ACCESSION		homo sapiens cDNA FLJ40170 fls. clone TEST12016848, weakly similar	
VERSION		AK097489	
KEYWORDS		AK097489.1 GI:21757270	
SOURCE		oligo capping; fls (full insert sequence).	
ORGANISM		Homo sapiens	
REFERENCE		Homo sapiens	
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
	1	Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwawaghi, T., Watanabe, M., Shiratori, A., Sudo, H., Hosori, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuma, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirao, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Houchi, T., Kusano, J., Kanohori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mueshino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsumura, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kunigata, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tachibana, T., Tanigami, A., Fujitani, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiroo, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kakumari, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyama, M., Hara, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, K., Nishikawa, T., Kimura, K., Yamashita, H., Kawai-Hiro, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuno, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Watanabe, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Matsuno, Y., Nagai, T. and Sugano, S. Iwagaki, T.	
TITLE		Complete sequencing and characterization of 21,243 full-length human cDNAs	
JOURNAL		Nat. Genet. 36 (1), 40-45 (2004)	
PUBMED		14702039	
REFERENCE		2	
AUTHORS		Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K., Arita, M., Mueshino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hiro, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuno, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Watanabe, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Matsuno, Y., Nagai, T. and Sugano, S. Iwagaki, T.	
TITLE		NEDO human cDNA sequencing project	

REFERENCE	LOCUS	DEFINITION	ACCESSION	KEYWORDS	ORGANISM	COMMENT
JOURNAL						Unpublished
REFERENCE						3 (bases 1 to 2534)
AUTHORS						Isoqai,T. and Yamamoto,J.
TITLE						Direct Submission
JOURNAL						Submitted (04-JUL-2002) Takao Isoqai, FwJ Project (HRI Team) ; 2-6-7
						Kazuo-Kamatari, Kitarazu, Chiba 292-0812, Japan
						(E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
						NEO human cDNA sequencing project supported by Ministry of
						Economy, Trade and Industry of Japan; cDNA full insert sequencing:
						Research Association for Biotechnology (RAB) ; cDNA library
						construction: Heiichi Research Institute (HRI) (supported by Japan
						Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
						HRI, and Biotechnology Center, National Institute of Technology and
						Evaluation; clone selection for full insert sequencing: HRI and
						RAB; annotation: HRI and RAB.
FEATURES						Location/Qualifiers
source						1..2534
						/organism="Homo sapiens"
						/mol_type="mRNA"
						/db_xref="taxon:9606"
						/clone="TEST12016848"
						/issue_type="cDNA"
						/clone_id="TEST12"
						/note="Cloning vector: pME18SFU3"
ORIGIN						
Query Match						100.0%; Score 401; DB 9; Length 2534;
Best Local Similarity						100.0%; Pred. No. 7e-71;
Matches						401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GAAGACCACTGGTGGCGCCAGGAGACCTGAGCTGCTGTGACTGTCAACGGCG	60			
DB	808	GAAACCACTGGTGGCGCCAGGAGACCTGAGCTGCTGTGACTGTCAACGGCG	867			
QY	61	GGAAGCGCCGCTGCACTGGCTGAAGACAGAAAGCGCCTCGCAAGACAGCAAGTAT	120			
DB	868	GGAAGCGCCGCTGCACTGGCTGAAGACAGAAAGCGCCTCGCAAGACAGCAAGTAT	927			
QY	121	GTGCTTCGACGAGGACAGTGGCCATGCTGTCTATCCGCGGAGCTTCTCTCAAGACCG	180			
DB	928	GTGCTTCGACGAGGACAGTGGCCATGCTGTCTATCCGCGGAGCTTCTCTCAAGACCG	987			
QY	181	GCGGACATCACTGTGAGTGGAGGCTTCCAAAGACACAGCCAGCTTCATGTGAGAA	240			
DB	988	GCGGACATCACTGTGAGTGGAGGCTTCCAAAGACACAGCCAGCTTCATGTGAGAA	1047			
QY	241	AAAGCAATGCTTCCACAGAGACCTGACCAATTCGACGTGGAGAGAAAGGACACACT	300			
DB	1048	AAAGCAATGCTTCCACAGAGACCTGACCAATTCGACGTGGAGAGAAAGGACACACT	1107			
QY	301	GTGCTTCAGTCAAGACAGGACACCCCGGAGCACAAGTGAAGCTGGCGCAAGGCTCTTG	360			
DB	1108	GTGCTTCAGTCAAGACAGGACACCCCGGAGCACAAGTGAAGCTGGCGCAAGGCTCTTG	1167			
QY	361	GAGCTACGGGCTCAGGAGACACACCGCCAGCGAGAGGG	401			
DB	1168	GAGCTACGGGCTCAGGAGACACACCGCCAGCGAGAGGG	1208			
RESULT 3						
BC043916		2761 bp mRNA linear PRI 23-JAN-2003				
LOCUS		Homo sapiens, similar to titin, clone IMAGE:3845393, mRNA, partial				
DEFINITION		cde.				
ACCESSION		BC043916				
KEYWORDS		BC043916.1 GI:27881999				
SOURCE		Homo sapiens (human)				
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		1 (bases 1 to 2761)				
		Strauberg,R.				

Db 8751 GCGGAGTACACGTGTGAGGTGGAGGCTTCAGAGACACAGCCCTTCATGTGAGAGAA 8810
Qy 241 AAAGCAAACTGCTTACAGAGAGAGTGAACCAATTCAGAGTGGAGAGAAAGGACAGCT 300
Db 8811 AAAGCAAACTGCTTACAGAGAGAGTGAACCAATTCAGAGTGGAGAGAAAGGACAGCT 8870
Qy 301 GTGTTACAGTGCAGAGACGAGAGACCCCGGCGGACAGTGAACCTGGCGCAAGGGCCTCTTG 360
Db 8871 GTGTTACAGTGCAGAGACGAGAGACCCCGGCGGACAGTGAACCTGGCGCAAGGGCCTCTTG 8930
Qy 361 GAGCTACGGGCTCTCAGGAGAGACCAAGCCAGCCAGAGAGG 401
Db 8931 GAGCTACGGGCTCTCAGGAGAGACCAAGCCAGCCAGAGAGG 8971

RESULT 7
LOCUS CQ445545 1005 bp DNA linear PAT 30-JAN-2004
DEFINITION Sequence 11305 from Patent WO0192523.
ACCESSION CQ445545
VERSION CQ445545.1 GI:41417003
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
AUTHORS Shinketo, R. A. and Leach, M. D.
TITLE Human polynucleotides and polypeptides encoded thereby
JOURNAL Patent: WO 0192523-A 11305 06-DEC-2001;
Curagen Corporation (US)
LOCATION/Qualifiers
FEATURES
source 1. 1005
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 98.0%; Score 393; DB 6; Length 1005;
Best Local Similarity 98.8%; Pred. No. 3.3e-69;
Matches 396; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GAAGACAGAGTGGTGGCGCCAGAGAGACGTGAGTTCGCTGTGAGCTGTACAGGCGG 60
Db 358 GAAGACAGAGTGGTGGCGCCAGAGAGACGTGAGTTCGCTGTGAGCTGTACAGGCGG 417
Qy 61 GGAACGCCCTGCTCACTGCTGAAGACAGAGAGCCATTCGCAAGAGCCAGAGTATGAT 120
Db 418 GGAACGCCCTGCTCACTGCTGAAGACAGAGAGCCATTCGCAAGAGCCAGAGTATGAT 477
Qy 121 GTGCTGTGAGAGGAGACGATGCGTGCATCCCGGGGCGCTCGCTCAAGAGACGG 180
Db 478 GTGCTGTGAGAGGAGACGATGCGTGCATCCCGGGGCGCTCGCTCAAGAGACGG 537
Qy 181 GCGGAGTACACGTGTGAGAGCTTCCAGAGACACAGAGCTTCATGTGAGAGAA 240
Db 538 GCGGAGTACACGTGTGAGAGCTTCCAGAGACACAGAGCTTCATGTGAGAGAA 597
Qy 241 AAAGCAAACTGCTTACAGAGAGAGTGAACCAATTCAGAGTGGAGAGAAAGGACAGCT 300
Db 598 AAAGCAAACTGCTTACAGAGAGAGTGAACCAATTCAGAGTGGAGAGAAAGGACAGCT 657
Qy 301 GTGTTACAGTGCAGAGACGAGAGACCCCGGCGGACAGTGAACCTGGCGCAAGGGCCTCTTG 360
Db 658 GTGTTACAGTGCAGAGACGAGAGACCCCGGCGGACAGTGAACCTGGCGCAAGGGCCTCTTG 717
Qy 361 GAGCTACGGGCTCTCAGGAGAGACCAAGCCAGCCAGAGAGG 401
Db 718 GAGCTACGGGCTCTCAGGAGAGACCAAGCCAGCCAGAGAGG 758

LOCUS HSA314905 34667 bp DNA linear PRI 21-MAY-2002
DEFINITION Homo sapiens partial OBSCN gene for obscurin, exons 31-56.
ACCESSION AJ314905
VERSION AJ314905.1 GI:21104337
KEYWORDS OBSCN gene; obscurin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
AUTHORS Young, P., Ehler, E. and Gautel, M.
TITLE Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 34667)
TITLE Direct Submission
Gautel, M. S.
JOURNAL Submitted (22-MAY-2001) Gautel, M. S., Physiologische Biochemie,
Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse
11, Dortmund, 44227, GERMANY
LOCATION/Qualifiers
FEATURES
source 1. 34667
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

gene
mRNA
CDS
join(247..519,1332..1598,1817..2082,4388..4664,5087..5353,6008..6274,6375..6638,6939..7202,10212..10475,10837..11100,11667..11930,12466..12729,13094..13357,16768..17031,17573..17836,18196..18459,21989..22252,22617..22880,23476..23700,24686..24952,25215..25481,25573..25846,26423..26697,27413..27609,27798..27873,34161..34433)
/gene="OBSCN"
/protein="Obscurin"
/product="Obscurin"
/protein_id="CAC85752.1"
/db_xref="GI:21104338"
/translation="LHVGITRRLTMEVLGEESCSFECVLSHESASDPAMTVGKTV GSSRFQATROGRKYILVREAPSDADGEVSVYGLTSKSLIVREBPAAIIRPLED OMVA RGEDEVRECELSRAGTPVHMLKORKAIRKOKYDVGEGTAMLVIRGASIKDA GEYTCVEYASKSTSLHYEERKANFTBELTTLQYBEKNTAVFTCTHPATPYVRKGL LLEIRASGKHQPSOGLTLRLTISALEAUSDYTCDDIGQNSRQQLVQGRVHITL ELDLEDVDVQEGSSATFRCPRISSPANYEPVHMLDPTPLANLEINLEIDAPGCVHITL QOLAKDSGTIFYEAGDQASALALVTREPSFSELDNDATTEGDLTLVETSCVDI PVCTKQDKTLIRYEGARCOLSHEGRALQILGATLQDSGRYKCEAGSSISVVA RPYVFOEALQLEVLREGAATLRCLVSVAPVPCVKNVLRQDQSLROEAMLE LVRNLRPODSGRYSCSGDDOTTSATLTVALTPQPIFLKANKEATBATNTLRLELS KAAPVEMKSGSETLRDGRYCLRDGACELQIQGLAVNDAAEYSCVGEERTSLSLT IRPFAHITGLRHQESTIGATATLRCELSSAAVEMKSGESLRDGRHSLRDQVAG CELQICGLAVADAGEYSCVGEERTSATLTVALPAKTEBGLRNEAVEGATLMAWCE LSKYAPVEMKSGPENLRDGRYILROEBRTCELOICGLAMADAGEYLCVCOERTSAT LTITLAPARFIEDVNOBARGATVALQCELSAAPVEMKSGSETLRDGRHSLRDQ TKECLOIRGLAMADAGEYSCVCOERTSATLTVALPAKTEBGLRNEAVEGATLVALR CELSKMAPEVEMKSGHETLRDGRHSLRDQKARCELQIRGLVABDAGEYLCMGKERTS AMLTVPAMPKSFIEGLRNEATEBDATLWCELSAAVEMKSGHETLRDGRHSLRDQ DSGRCELOIRGLAVADAGEYSCVCOERTSATLTVALPAFIEDVNOBARGATVAL LQCELSKRAPEVEMKSGSETLRDGRYSLQDGTCELOIHLSVADQYEVSCVCOER TSATLTVALPARFIEDVNOBARGATVALQCELSKRAPEVEMKSGSETLRDGRYSL QDGTCELOIHLSVADQYEVSCVCOERTSATLTVALPARFIEDVNOBARGATVALR ATIQCELSKRAPEVEMKSGLELRDGRKTSLSQDQAVCELOIHGLAMADNGYSCVCGQ ERTSATLTVALPARFIEDVNOBARGATVALQCELSKRAPEVEMKSGPNTLKDDRY SLKQDGTSCSELQIRGLVADAGEYSCICEOERTSATLTVALPARFIEDVNHTEG

CDS

/gene="OBSCN"
 72..19934
 /gene="OBSCN"
 /function="t1c1n binding"
 /codon_start=1
 /evidence=experimental
 /product="obscutrin"
 /protein_id="CAC44768.1"
 /db_xref="GI:15026974"
 /db_xref="GOA:Q96AA2"
 /db_xref="UniProt/T-EMBL:Q96AA2"
 /translation="MDQPSGAPRRLTRKAVVSVGKDATISQIVGNPPTQVSW
 KDQPTPAAGRFLADQDGLRYLTLDLADGQYVCRAKNAIGAFVAVGQVAVAE
 AAQABQPHFLNPTSLRREGESEATRCRVGSPRAVWSMDKGRLEGPDPBRVAV
 EELGEASALRIARAAPRDGTYEVRKNEPLGAASAAALVDDADLTSPGSTRA
 LLAHLORREARMAEGAPASPSTGRTCTVTEGKARLSCTYGSPEKPTVXKQGL
 LVEGRHVVYEDAGENFVLKIFCKSDRGLYCTASNLVQOTSSVTVVVRPAPV
 FKRLDLEVEREKESATPLCEVPDSTEAMFEEBRLNASKAKYIEEBSTERLTV
 NVSADDAVYICETPSGRVVALAQQNLRLKLPKRTARVGDIMPCVELAVVGP
 VMLRNOEEVAVAGRAVISAEGTRHTLTISQCLBDVQVAFMAGDCQSTRCVSAP
 RKPLDPLVADVAVKAMESSVILMSPPRGERPVTIDGLVEKKLGTVMYRCHBA
 EMVATPELTVADVAEAGNFOFRVSAI NSFGOSPYLEPPTVHLAPKLAVRTPYKAAQ
 VEGEVTFESVDTLVAAGEMFLDGLAKSSVYEIHCDDTRHTLTIRVPASIHGQL
 KEVANGIESIRMEVRAARGLTANKPRAAAREVTLRLHBEAQLALSDQAAAVTL
 KDGRITSPGKTYVQASGRVTLVADVADADGLYECVSRGRLAYQLSVQGLAVTL
 HKWAGSCVDAVAGAPAOEFCETSEAHVHVHVKDMLHSGERFLQBDVGRRLV
 AATVTDQDEGTVSCRVBDSVDFRLRVESEKVFVFAKOLARLLOAEAGASATLSCEV
 AAQOTVMTYKDKGKLSSSKVCMEATGCTRLVVOAGQADAGEYSCAAGRLSFH
 LDYKEPRVFAKDOVAHSEVOAAGANATLSCEVQAQAEVWVYHDKGLSSIKYHV
 BAKGCRRLVVOAGKTDAGDYSCEARGORVSRHLITPRKMFPAKESVHNSVOEA
 GASAMLSCVAAQOTVMTYKDKKLSSSKVCMEATGCTRLVLPQAKADAGEYSC
 BAKGQDVSFHLHITTEKGVFAKESVHNSVOEAAGVTAALSCVAPQTEVMYKDGK
 KLSSSKVCMEATGCTRLVVOAGQADAGEYSCAAGRLSFHLDYKEPRVFAKDO
 LVHNEVRETAGASATLSCEVAAQOTVMTYKDKKLSSSKVCMEATGCTRLVVOA
 GQADAGEYSCAAGRLSFHLDYKEPRVFAKDOVAHSEVOAAGANATLSCEVQA
 TMTYKDKKLSSSKVCMEATGCTRLVVOAGQADAGEYSCAAGRLSFHLDYKEPRV
 EKVVFAKDOVAHSEVOAAGANATLSCEVQAQAEVWVYHDKGLSSIKYHV
 CTRRLVVOAGQADAGEYSCAAGRLSFHLDYKEPRVFAKDOVAHSEVOAAGANAT
 TLSCEVAAQOTVMTYKDKKLSSSKVCMEATGCTRLVLPQAKADAGEYSCAAG
 OSLFHLHAEKPRVFAKDOVAHSEVOAAGANATLSCEVQAQAEVWVYHDKGLSS
 SKVREVAAGCTRLVVOAGQADAGEYSCAAGRLSFHLDYKEPRVFAKDOVAHSE
 PLVKEHEDILTLATLTPSAATVMTYKDKGVLEIRSKHETASQDHTLTVAGQVL
 DSATYSCRVAAGDOPVQVETVAAKFCGLLEPVGELGTYTLACBISPAACAVYK
 CQNTOPRVKRPQMAAEGPVRLTYGLRAEDAGEYSCSRDHTSAQITVSPVYK
 PMSGLTVAAEGGEATFQCVVPSDVAWVFRDALQPSKFAISQSGHSLTIS
 DLVLDAQOITVEABGASSALRVEAEVTLFKKLEPQTEVERSSVLEVELTTPM
 ELRTNRNATALAPGKNVEIHAAGARRLVANVGRADRGFCETPDDTKOAKLVEM
 ROVRLVGLQAVAREOGGTATMEVOLSHADVDSMTROGLRQOGPTCHLAVRGMMT
 LLSGLRPEDSGLAVPKAEGVTSARLVYTEL PVSFSPRLDQVYVTEKKTILBCELS
 RENDVRLKDOVEIRAGKTAIAAQACRSLLTYRCEPADQVYVCAHDAQSSAV
 KYQKRTYTLIRKVLAEADGETIQVAENASRAQLRYELPVTIAPRLDKIAMEKHR
 GYLECQVSAQVAMFKGSOBLQPGKYEIVSDGLYRKLISDVHADDEPTYDAG
 DVTEQGEVTEHMLRRTCTMTGTVMEPAWPFCECETSPSPRPMKLLDKTYVAC
 GNVLEOEGTVHMLRRTCTMTGTVMEPAWPFCECETSPSPRPMKLLDKTYVAC
 RELQSVLSCDPRPAKAVOMYKDDTPEPSKPKMSLEGQAAELRILRLMPADQVY
 RQAGSSASTTEVTEARVYTTGFLQDAEATEBEMASFCELSHEDVEVMSNGMP
 LYNDPFHEISHGRHTVLKSTIQADAGVYASSLKYSTARLEVRKVPVFKALD
 DLSAEERGTALQCEVSDPEARVAVVRKQVQGPDKDPLHTAGRLVAVHSPED
 AGLYTCYHGESETRARVVDHLVGIITKRLKMEVLEBESGSEFVLSHESASPMY
 TVGKTVGSSSRFOATROGRKYLIVRAAEKADAGEYVSRGLSKSLIVRRPAA
 IIKPLEDOMVAGDEVELRCFLSRAGTGVHMLKDRKARKSQKIDVYCEGTMATVIR
 GASIKLEADGETVCEVASKSTASLHYEBEACFTEBLTNLYQVEKGTATCTHHPAA
 TVTRKGLLEJASGRKQPSQEGTLTLTILALEKADVDTYCTIGQAQSTQAQLVQV
 RRVHILIEDLEVDVQESSATRCRISIPANVEPVMFJDKPLPLANENIEIDAQGY
 HVLTLRQALNDGSGTIFEADQORASALRYEKESVSRDLPTANTIEBGLTLVCE
 TSTCDIPMCWTQDKGTLTLAGSARCOLSHEGHAOLITGATLQDSGRKYKCEGACSS
 IIVRYHARVPRFOBALKDLLEVLBGAAITRCTLSSVLAAPKVCYGNVILRPDCKSLQ
 EGAMTELIVRMLRPDGSRYSCSPGDTTSTLTYTALPAQFIRLARKETKTECATO
 LRCELSKTAIYVWRKGSITLRDGDYCKRQDQAMCELOIRELANWDALEYSVCGEER
 TSASLITRPMFAHIGRLRHQBSLEGATPALTRKESKAPREKMGESLSDGRHSL
 RQDAVCELOIQGLAVADAGEYSCVCGERTSATLTYVALALCKTEGIRNEAVTEG
 AMLWCELSKAPVEMRKGPENLRDGRITLROEGTRCLOIQGLMADAGEYLVCCO
 ERTSATILIRALPARFIEDVKNQARBAATVLOCELNSAPVEMRKGSITLRDGRY

ORIGIN

Query Match	100.0%; Score 401; DB 9; Length 20435;
Best Local Similarity	100.0%; Pred. No. 5e-71;
Matches	401; Mismatches 0; Indels 0; Gaps 0;
QY	1 GAAGACCAAGTGGGGTGGCGCCAGAGAGACGTTGAGAGCTGGCTGTGAGCTGTCAACGGCG 60
Db	8571 GAAGACCAAGTGGGGTGGCGCCAGAGAGAGACGTTGAGAGCTGTGAGCTGTCAACGGCG 8630
QY	61 GGAACGCGCGTGGACCTGGCTGGAAGGACAGGAAGCCATCCGCAAGAGCCCAAGAGTATGAT 120
Db	8631 GGAACGCGCGTGGACCTGGCTGGAAGGACAGGAAGCCATCCGCAAGAGCCCAAGAGTATGAT 8690
QY	121 GTGGTCTGCAAGGAGCAAGATGACCATGCTGTGATCCGCGAGGCTCTGCTCAAGACGCG 180
Db	8691 GTGGTCTGCAAGGAGCAAGATGACCATGCTGTGATCCGCGAGGCTCTGCTCAAGACGCG 8750
QY	181 GCGGAGTACACGTTGTGAGGCTTCCAGAGACACAGCCAGCTTCCATGTTGGAAGA 240

SLRQDGTKELOIRGLAMADNGEYSCVCGERTSATLTYVALPIKFTTEGLRMEARETEG
 ATAVRCELSKAPVEMRKGPENLRDGRITLROEGTRCLOIQGLMADAGEYLVCCO
 ERTSATILIRALPARFIEDVKNQARBAATVLOCELNSAPVEMRKGSITLRDGRY
 GERTSATILIRALPARFIEDVKNQARBAATVLOCELNSAPVEMRKGSITLRDGRY
 RSHLSDGSRCELOIRGLAVDAGEYSCVCGERTSATLTYVALPARFIEDVKNQAR
 BGATAVLOCESKAPVEMRKGSITLRDGRITLROEGTRCLOIQGLMADAGEYSCV
 CGERTSATILIRAPVEMRKGSITLRDGRITLROEGTRCLOIQGLMADAGEYSCV
 VCGERTSATILIRAPVEMRKGSITLRDGRITLROEGTRCLOIQGLMADAGEYSCV
 ANGRREPRLOGCTGTAELVLODLQREDTGTGCTGCGATATVLAAPRFLRELOD
 EYDREGTANLCCELSRAGAVEMRKGSITLRDGRITLROEGTRCLOIQGLMADAGEY
 DYTCDGHTQSNASLSVYPRPKRTLRQSLQEQETGDLARLCCQLSDSAGVAVOMLK
 BEVELHAGPKTSMRQATRELLIHOLEKATDGEYACVYQDQKTAASLRDEVTIY
 RCLVDAEVLADDEVESEVSRAGATGVOMCLOGLPLQSNTEVTEARVADGRITHTRLK
 GTTPBDAGVSEFHLNHAASQOLTVRAEVEITLEPQVLOVQSEGDASFOCQLSRASG
 QEARWALGGVPLQANEMNDIYVEOGLTHLTHKTYLIDAGTVSPHVGCSSEADQV
 TAKNTVAGLEVEALREGGALFEQLSQEPVAAHTMLLDEPTEVSNAPVPEVNG
 LRLHLLKNLRPODSCTPTFLAGDMTSAFLTVRAETMLIEPLKMAVRAQAQRFT
 CTLSAIVYVGEBSWRTINQAQVQPDSDMTVADSGHQLKLSAPPHAGETVACRD
 AVASARLTVLGLIPDEDEAVAHSHSVTLISMAPMDDGGGLGCVREVEAGATQ
 WMLCHELVGPECVVDGLAPGTYAFRFAVAVPGVAGRPVHLPOTVRLAEPKVPVQ
 PSAPSRQVAAAGEVDLSLEVAEAGEVIMKGMERIQGGRFEVVSQROOMLYIKG
 FLAEDQGEYHGLAQGSI CPAATFOVALSPASVDEARPOSLPREAAGEGLHLIWEA
 LARKRMRPEPTLDSISLTPREDRSQRLPOBAEVAPDISEGSTADELARTGDAL
 SHTSDDSRKAGTSLVLYLKQAGRGTSPLASKVGAAPSVFQOQDEPLAARVAP
 LGDSTKOLGDPMDKAUVKIQAAPFKGVKEMKQDGPMSHFTGDEKQVGLAR
 LECVVASADVARAMLKXGVELITDGRHHHIDQIGTCSLLIAGLDRADACYTCQVS
 NKFGQVTHSACVAVVSGSESEASSGGELDAPRAARLHRLPRTTSPARVSDLEF
 LSADGPAREPE
 VBOGPRVEMCISKETPAPVYPPRPLBSLTSDAPVFLTELQNEVDGTPVSDVC
 VTQGPMSVRRFKKGLLEBDHMTINDDQGHQOLITAVPDMGVYRCLANSWG
 VSTYAEIRVBLTSTDYTDADATSESSYFAQOQLSRHEGDESTDEQLQVVE
 BLRDLQVAPGTRLAKFOLKGVYAPAPRLYFVKQDPLASAHIMTQKILHTEIIS
 VTRBDSGGYAAVYINAMGAAYSAARLVGRGDEEBEPASVDHQLVAPRMLBETPK
 KYKSGSITFVKTEGEPVPTVHMLREAEAGVMIQDPTGVHQLVAPRMOHSVLID
 VGRHOQGYTCOISANAGQOLCASLHVSGLPKYREBOKYKALISPTLOCTQTAISA
 QGPOTASPADQCGORKEERPLAAKALGHLISLAVGTSEFLQKLSQITTEMASCTQIA
 KLQVPGSDSDSDSTPSAPRHHGSRSSSIQESSSESDADAGEIPDYLVAVDLY
 PLGABQDITLREQVYVLDAAHPLVMTKTKSPSSPSQGWVSAVYLDRLKLP
 EBMGAELPEFGESEVDEYKARLSYIOELISEBOAFLVELOCTLOSHHLOHLRCPH
 VPIVAGQKAVIFPNVRDIIGRFHSFLOELQOQCTDDVAMCFIKNOAAPFOXYELLK
 GRVQESVSVETALIOEPKPKYKAEALLAGDSDPPDLMYCELOPBAVRVORVOLLK
 ELINKKARNRQNCALLQAVYALPQRAKNTKATVLMENYPTDQALBEPHQGH
 IWEKAGDARPRKGNHRHFTLFNNHLLITCERKDSSTDVYSVFRMMKLSITDLND
 QVEDDGAFFEWQREBDSVKRYLLQAAHTVALIKSSWVEIKQIQORLAPVWRPDEF
 ELADCTAELGTYVLAQVGTCPKPVLSVWPRVNVVLASPFVEGDADPTCTIBGA
 LTVGDSQYQMFASAAAGNCSTGLKILVQVPRVNVVLASPFVEGDADPTCTIBGA
 PPIRYMKQGALLTGNKFOITSEPSGLITVYRAASKEDGLYCELYVNRGSGAR
 ASAEIRIOSMPLQAOBQCHREOLYAEVTEOETKVPYKXTYIIEETITTVYKSGOR
 RSPKSSRSRPSRSASPLRGLIADPLTLDPGAGQRPTEAEFGQRPVPTLVYTA
 EAHBPALPGLSGPQPKAVEVEETIYKVKVMGPGVSPTEVPRSSGHLFTLPGATP
 GGDPSNNNSNNKLLAOEAMAOAGTAMVGRMPLVFRVARSVMAASGMSLEBETM
 BEAGEEBEGEDGAFVTEESODTHSLGDRDKLITNNGRMLTLADIEDYVGEGETFHC
 GGPQGPADDPDPCGVSVIOREIGEPVGOVVLISVGHALGMDPLGLFRPEPGASPP
 GPQVRSLEGSFLLREAPARPVGSAPMTGSCFTRIRASDSGGSFTTELSTQTVNFG
 TVGRTVTLHLCPRDDBEAQP"

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

FEATURES	Location/Qualifiers
SOURCE	1. .135964

NYSTATNO

	Query Match	59.8%	Score	239.8	DB	9	Length	135964
	Best Local Similarity	99.2%	Pred.	No. 1e-38				
	Matches	241	Conservative	0	Mismatches	2	Indels	0
	Gaps	0						
Qy	1	GAAGACCAATGCTGGTGCAGCGGAGAGGACTGTGAGCTCCCTGTGACTGTTCACAGGCGG	60					

RESULT 10	AC026657	LOCUS	DEFINITION
AC026657	164766 bp	DNA	linear HTG 01-SEP-2000
AC026657	Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT		
AC026657	SEQUENCE, 31 unordered pieces.		
AC026657	AC026657		
AC026657.4	GI:9958202		
HTG; HTGS PHASE1; HTGS_DRAFT.			
Homo sapiens (human)			

----- Genome Center -----

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index\_shtml
----- Project Information -----
Center project name: H_NH0245P10
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid, 0%
Chemistry: Dye-primer ET, 100% of reads
Chemistry: Dye-terminator Big Dye, 0% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 139884 bases at least Q40
Consensus quality: 147686 bases at least Q30
Consensus quality: 151469 bases at least Q20
Insert size: 169000; agarose-fp
Insert size: 161074; sum-of-ctrls
Quality coverage: 3.60 in Q20 bases; agarose-fp
Quality coverage: 3.92 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1124: contig of 1124 bp in length
* 1125 1224: gap of unknown length

```

ATAVLQCELSKAAPEWKRKSEITLRDGRYSLRQDTRCELQIRGLAVEDTGEYLQVC
GQERTSATLTVRALPARFIDNMTNOEARBEGATILHCELSKVAPEWKRKPEITLRDQ
RHSIRQDGTCELRIGLSVADAGEYSCVGQRTSATLTIRATGATMLQCELSK
VAPEWKRKPEITLRDGRYNLRQDGTCELRIGLSVADAGEYSCVGQRTSATLTIR
KAPQVPREPLQSLQAEBSSTATLQCELSPTATVYWSKGLQLOQANGREPRELQOCT
AEVLQDLOREDTGEYTCCTGSOATSTTTVTYTAAPRFLRELOHQVEDSGTNAHLCCE
LSRAGASVEMRKSLQLFPCAKYQWVODGAABLVRGVEQEDAGDYTCPTGTQXMA
SLSVRVRPFRKTRLOSLQEBETDIARLCCQLSDASGAVVQWLKEGVELHAGPYEM
RSQGATRELLHQLEAKDTGEYACVTGGQTAASLRVTEPEVTIRGLVADAEVDE
DVEFSCVSRAQATGVQWCTQGLPLQSNVEYEAVRDGRHITRLKGVTPEDAGTVP
HLGNMSSAQLTVRAPEVTILREPLQVQVLSQGDASRQCELSRASQGEARMAIGVPL
QANEMNDITVEOGTGLHLTLAKVTLQEDAGTSHVGTCSSEALQKTKAKTVKRGLEN
VEALREGGALFECQLSQPEYAAHTWLLDDEPHTSNAEYVFPENGLRHLLKKNLRP
QBSCRVTFIAGDVTSAFLTVR"
247. .519
/gene="OBSCN"
/number=31
520. .1331
/gene="OBSCN"
/number=31
1332. .1598
/gene="OBSCN"
/number=32
1599. .1816
/gene="OBSCN"
/number=32
1817. .2082
/gene="OBSCN"
/number=33
2083. .4387
/gene="OBSCN"
/number=33
4388. .4664
/gene="OBSCN"
/number=34
4665. .5086
/gene="OBSCN"
/number=34
5087. .5353
/gene="OBSCN"
/number=35
5354. .6007
/gene="OBSCN"
/number=35
6008. .6274
/gene="OBSCN"
/number=36
6275. .6374
/gene="OBSCN"
/number=36
6375. .6638
/gene="OBSCN"
/number=37
6639. .6938
/gene="OBSCN"
/number=37
6939. .7202
/gene="OBSCN"
/number=38
7203. .10211
/gene="OBSCN"
/number=38
10212. .10475
/gene="OBSCN"
/number=39
10476. .10836
/gene="OBSCN"
/number=39
10837. .11100
/gene="OBSCN"
/number=40
11101. .11666
/gene="OBSCN"

	Query Match	59.8%; Score 239.8; DB 9; Length 34667;
	Best Local Similarity	99.2%; Pred. No. 1.3e-38;
	Matches 241; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY	1 GAAAGCCAGTGGGTGGGCCAGGAGAGAGCTGAGCTGCTGTGAGCTGTACAGGGCG 60	/number=40 11667. .11930 /gene="OBSCN"
Db	1361 GAAGACCAGTGGGTGGGCCAGGAGAGAGCTGAGCTGCTGTGAGCTGTACAGGGCG 1420	/number=41 11931. .12465 /gene="OBSCN"
QY	61 GGAACGCCGTGACCTGGCTGAAGGACAGGAGGCCATCCGCAAGGCCAGAGTATGAT 120	/number=41 12466. .12729 /gene="OBSCN"
Db	1421 GGAACGCCGTGACCTGGCTGAAGGACAGGAGGCCATCCGCAAGGCCAGAGTATGAT 1480	/number=42 12730. .13093 /gene="OBSCN"
QY	121 GTGCTGCGAGGCGCAGATGCGCATGCTGTCATCCGCGGCGCTGCTCAAGAGCGG 180	/number=42 13094. .13357 /gene="OBSCN"
Db	1481 GTGCTGCGAGGCGCAGATGCGCATGCTGTCATCCGCGGCGCTGCTCAAGAGCGG 1540	/number=43 13358. .16767 /gene="OBSCN"
QY	181 GCGGAGTACAGCTGTGAGGTGAGGCTTCCAGAGGACAGCCAGCTTCATGTGAGGAA 240	/number=43 16768. .17031 /gene="OBSCN"
		/number=44 17032. .17572 /gene="OBSCN"
		/number=44 17573. .17836 /gene="OBSCN"
		/number=45 17837. .18195 /gene="OBSCN"
		/number=45 18196. .18459 /gene="OBSCN"
		/number=46 18460. .21988 /gene="OBSCN"
		/number=46 21989. .22252 /gene="OBSCN"
		/number=47 22253. .22616 /gene="OBSCN"
		/number=47 22617. .22880 /gene="OBSCN"
		/number=48 22881. .23475 /gene="OBSCN"
		/number=48 23476. .23700 /gene="OBSCN"
		/number=49 23701. .24685 /gene="OBSCN"

TITLE
JOURNAL

REMARK
COMMENT

Direct Submission
Submitted (10-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapps@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAL Plate: 44 Row: c Column: 8.
Location/Qualifiers
1. .2761
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3845393"
/tissue_type="placenta, choriocarcinoma"
/clone_lib="NIH MGC 21"
/lab_host="DH10B-R"
/note="Vector: pOT7"
<1. .2306
/codon_start=3
/product="Similar to titin"
/protein_id="AAH43916.1"
/db_xref="GI:27682000"
/translation="DNVIRQERTYLLIYRVLAEDAGETQFVAENAESRQLRVKE
LPVILRLPRDKIAEKHRLVLEQVSASQVRFKSGLOELQPGKPYELVDFGLYRK
LIISDVHAEDEDTYCDAGDVYKSAQFFVEEQSITVIRGLQDVMPAPAFWPCFETS
IPSVRPKWLIGTKVQAGNVGLEQGTVHRLMRTCTMTGTVHFTVSKRSAR
LWSDIPVLLPRLPEKTRGLQSVLSCDFRPAPKAVQWKDTPLSPEKPKWSLE
GMAELRLRLMPADAGVYRCQAGSAHSSTEVVEAREVTVGLQDAEATEGWASF
SCELSHEDEEWSLNGMPLVYDSFHEISHKRRHTLVLSIQRADAGIVRASSLKVS
TSARLEVRKPVVFLKALDDLSBERGTLAQCEVSDPEAHVWRKDGVLQSPDKDY
FLHTAGTGLVHDVSPEDAGLYTCHVGSSETRARVRVHDLHVGITKRLKTMVELEGE
SCSFEVLSHESASDPAMWTGGTGVSSSRFOATROGRKYILVVRAPSDAGEVVF
SVRGLTSAKSLIVRERPAIITKPLEDQWVAGEDVELRCELSTRAGTVPVHLKDKAIR
KSOYDVVCEGTAMLVIRGASLADAGEYTCVEASKSLHVEEKANCFTEBLTNL
QVEKGTAVFCKTEHPAATVWRKGLLELRASGHQPSQEGTLRLITLSALEKADSD
TYTCDIGQAQSRALLVOGEAAKCARA"

FEATURES
source
1. .2761
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3845393"
/tissue_type="placenta, choriocarcinoma"
/clone_lib="NIH MGC 21"
/lab_host="DH10B-R"
/note="Vector: pOT7"
<1. .2306
/codon_start=3
/product="Similar to titin"
/protein_id="AAH43916.1"
/db_xref="GI:27682000"
/translation="DNVIRQERTYLLIYRVLAEDAGETQFVAENAESRQLRVKE
LPVILRLPRDKIAEKHRLVLEQVSASQVRFKSGLOELQPGKPYELVDFGLYRK
LIISDVHAEDEDTYCDAGDVYKSAQFFVEEQSITVIRGLQDVMPAPAFWPCFETS
IPSVRPKWLIGTKVQAGNVGLEQGTVHRLMRTCTMTGTVHFTVSKRSAR
LWSDIPVLLPRLPEKTRGLQSVLSCDFRPAPKAVQWKDTPLSPEKPKWSLE
GMAELRLRLMPADAGVYRCQAGSAHSSTEVVEAREVTVGLQDAEATEGWASF
SCELSHEDEEWSLNGMPLVYDSFHEISHKRRHTLVLSIQRADAGIVRASSLKVS
TSARLEVRKPVVFLKALDDLSBERGTLAQCEVSDPEAHVWRKDGVLQSPDKDY
FLHTAGTGLVHDVSPEDAGLYTCHVGSSETRARVRVHDLHVGITKRLKTMVELEGE
SCSFEVLSHESASDPAMWTGGTGVSSSRFOATROGRKYILVVRAPSDAGEVVF
SVRGLTSAKSLIVRERPAIITKPLEDQWVAGEDVELRCELSTRAGTVPVHLKDKAIR
KSOYDVVCEGTAMLVIRGASLADAGEYTCVEASKSLHVEEKANCFTEBLTNL
QVEKGTAVFCKTEHPAATVWRKGLLELRASGHQPSQEGTLRLITLSALEKADSD
TYTCDIGQAQSRALLVOGEAAKCARA"

CDS
100.0%; Score 401; DB 9; Length 2761;
Best Local Similarity 100.0%; Pred. No. 6.9e-71;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 6.9e-71;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GAAGACCAAGTGGTGGCCAGGGAGGACCTGGAGCTGCCTGTGAGCTGTCCACGGCG 60
|||||
1773 GAAGACCAAGTGGTGGCCAGGGAGGACCTGGAGCTGCCTGTGAGCTGTCCACGGCG 182
|||||
61 GGAAGCCCGTGCCTGGCTGAAGACGACGAGGCCATCCGACAGGACGACGATATGAT 120
|||||
1833 GGAAGCCCGTGCCTGGCTGAAGACGACGAGGCCATCCGACAGGACGACGATATGAT 1892
|||||
121 GTGCTCTGCGAGGCGACGATGCCATCTGCTCATCCGCGGGGCTCGCTCAAGGACGCG 180
|||||

1893 GTGCTCTGCGAGGCGACGATGCCATCTGCTCATCCGCGGGGCTCGCTCAAGGACGCG 1952
181 GCGCAGTACACGCTGTGAGTGGAGGCTTCCAAAGACACAGCCAGCCTCCATGTGGAAGA 240
1953 GCGCAGTACACGCTGTGAGTGGAGGCTTCCAAAGACACAGCCAGCCTCCATGTGGAAGA 2012
241 AAAGCAAACTGCTTCCACAGAGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACGCT 300
2013 AAAGCAAACTGCTTCCACAGAGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACGCT 2072
301 GTGTTACAGTGCAGAGAGGAGACCCCGGGGACACAGTGCCTGGCGCAAGGGCCTCTTG 360
2073 GTGTTACAGTGCAGAGAGGAGACCCCGGGGACACAGTGCCTGGCGCAAGGGCCTCTTG 2132
361 GAGCTACCGGGCTCAGGGAAGCACCAGCCAGCCAGGAGGG 401
2133 GAGCTACCGGGCTCAGGGAAGCACCAGCCAGCCAGGAGGG 2173

AB046776 Homo sapiens mRNA for KIAA1556 protein, partial cds.
AB046776 Homo sapiens
AB046776.1 GI:10047176
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nagase, T., Kikuno, R., Nakayama, M., Hirose, M. and Ohara, O.
Prediction of the coding sequences of unidentified human genes.
XVIII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
DNA Res. 7 (4), 273-281 (2000)
20450683
10997877
2 (bases 1 to 4789)
Ohara, O., Nagase, T. and Kikuno, R.
Direct Submission
Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail: cdnaif@kazusa.or.jp;
URL: <http://www.kazusa.or.jp/huge>, Tel: 81-438-52-3913,
Fax: 81-438-52-3914)
Location/Qualifiers
1. .4789
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fh18619"
/notes="vector: pBluescriptII SK plus"
1. .4789
/gene="KIAA1556"
/notes="KIAA1556"
/notes="Start codon is not identified."
/codon_start=3
/product="KIAA1556 protein"
/protein_id="BAB13382.1"
/db_xref="GI:10047177"
/translation="RARVRVHDLHVGITKRLKTMVELEGSCSFEVLSHESASDPAM
WTGGTGVSSSRFOATROGRKYILVVRAPSDAGEVVFVSRGLTSAKSLIVRERPA
AIIKPLEDQWVAGEDVELRCELSTRAGTVPVHLKDKAIRKSOYDVVCEGTAMLVIR
GASLADAGEYTCVEASKSLHVEEKANCFTEBLTNLQVEKGTAVFCKTEHPAATV
WRKGLLELRASGHQPSQEGTLRLITLSALEKADSDTYTCDIGQAQSRALLVOGEAA
KCARA"

FEATURES
source
1. .4789
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fh18619"
/notes="vector: pBluescriptII SK plus"
1. .4789
/gene="KIAA1556"
/notes="KIAA1556"
/notes="Start codon is not identified."
/codon_start=3
/product="KIAA1556 protein"
/protein_id="BAB13382.1"
/db_xref="GI:10047177"
/translation="RARVRVHDLHVGITKRLKTMVELEGSCSFEVLSHESASDPAM
WTGGTGVSSSRFOATROGRKYILVVRAPSDAGEVVFVSRGLTSAKSLIVRERPA
AIIKPLEDQWVAGEDVELRCELSTRAGTVPVHLKDKAIRKSOYDVVCEGTAMLVIR
GASLADAGEYTCVEASKSLHVEEKANCFTEBLTNLQVEKGTAVFCKTEHPAATV
WRKGLLELRASGHQPSQEGTLRLITLSALEKADSDTYTCDIGQAQSRALLVOGEAA
KCARA"

gene
CDS